Electric Vehicle Market Segmentation

Team Members:

1. Aniket Tidke: https://github.com/TidkeAniket/Feynn-Labs-Project-2-.git
2. Vikrant Arora: https://github.com/auroravicks/Feynn-Labs/blob/main/EV.ipynb
3. Kamlesh Prajapat: https://github.com/Kamleshp11/EV_Market_Segmentation
4. Pratim Maity: https://github.com/Pratim-72/Car-Data-Analysis
5. Ruturaj Patil :https://github.com/ruturaj-p/Data/tree/main

Fermi Estimation:

To better understand the market segmentation of the Electric Vehicle (EV) market in India, let's perform a Fermi estimation based on the given problem statement.

1. Estimated Total Potential Customers: Considering the population of India (approximately 1.3 billion), let's assume a certain percentage of individuals are potential customers for EVs based on factors such as affordability, awareness, and environmental consciousness. Assuming a conservative estimate of 10% of the population as potential customers, we can estimate around 130 million potential customers.

2. Seating Capacity Preference: Let's assume that customers in the EV market have varied seating capacity preferences. Based on general observations and market trends, we can estimate that around 40% of potential customers prefer 5-seater cars, while 30% prefer 4-seaters and the remaining 30% prefer 7-seaters. This estimation results in approximately 52 million potential customers for 5-seaters, 39 million for 4-seaters, and 39 million for 7-seaters.

3. Price Range Distribution: Assuming a range of EV prices in the market, let's estimate the distribution of potential customers across different price ranges. Based on market knowledge and analysis, we can estimate that approximately 25% of potential customers are interested in EVs priced below 5 lakhs, 40% in the range of 5-10 lakhs, and the remaining 35% in the range of 10-30 lakhs. This estimation results in approximately 32.5 million potential customers in the below 5 lakhs price range, 52 million in the 5-10 lakhs range, and 45.5 million in the 10-30 lakhs range.

4. Horsepower Preference: Assuming an equal distribution of customers across horsepower preferences, we can estimate that approximately 650 million potential customers in India seek EVs with horsepower above 200.

These Fermi estimates provide a broad understanding of the market segmentation for the EV market in India. Further analysis and data validation are necessary to refine these estimations and gain more accurate insights into the specific market segments.

Data Sources:

https://www.nber.org/research/data/cross-country-historical-adoption-technology
https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset
https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho
https://www.statista.com/topics/1487/automotive-industry/#topicOverview
https://drive.google.com/drive/folders/1yGp_ahJznyWBUFzjkR8EfRpmJnHuddqB
https://drive.google.com/file/d/1mshWg1hYSG5ifT3988eDCeGTWnSd8Z2n/view?usp=sharing

Data Preprocessing and Segment Extraction：

Aniket Tidke：

Data set is divided into two decades for comparison.

Year 2001 to 2010:

Table: Cars data from decade 2001 to 2010

No of seats	Total cars	Average of engine(CC)	Average of mileage(km/l)	Average of max_power(bhp)
4	60	784.533	16.75	36.908
5	984	1276.795	17.581	78.885
7	115	2427.93	12.862	110.995
8	70	2249.129	13.198	98.72
9	13	2283.923	12.845	92.3
10	7	2478	14.157	75.143
14	1	1948	10.71	90

Here are some insights from the provided data:

Seat Distribution: The majority of cars in the dataset have 5 seats, accounting for 78.72% of the total cars. This is followed by cars with 7 seats (9.2%), 8 seats (5.6%), 4 seats (4.8%), 9 seats (1.04%), 10 seats (0.56%), and finally, cars with 14 seats (0.08%). This indicates that cars with 5 seats are the most popular choice among consumers.

Engine Size and Seat Count: The average engine size, measured in CC, generally increases with the number of seats. Cars with 10 seats have the largest average engine size of 2478 CC, followed by cars with 9 seats (2283.923 CC), 8 seats (2249.129 CC), 7 seats (2427.93 CC), 14 seats (1948 CC), 5 seats (1276.795 CC), and finally, cars with 4 seats have the smallest average engine size of 784.533 CC. This suggests that larger vehicles with more seats tend to have more powerful engines.

Mileage and Seat Count: The average mileage, measured in km/l (kilometres per litre), varies across different seat counts. Cars with 5 seats have the highest average mileage of 17.581 km/l, followed by cars with 4 seats (16.75 km/l), 8 seats (13.198 km/l), 9 seats (12.845 km/l), 7 seats (12.862 km/l), 10 seats (14.157 km/l), and finally, cars with 14 seats have the lowest average mileage of 10.71 km/l. This indicates that cars with 5 seats tend to be more fuel-efficient compared to larger vehicles.

Power and Seat Count: The average maximum power output, measured in bhp (brake horsepower), also shows variation based on the number of seats. Cars with 7 seats have the highest average maximum power of 110.995 bhp, followed by cars with 8 seats (98.72 bhp), 14 seats (90 bhp), 9 seats (92.3 bhp), 10 seats (75.143 bhp), 5 seats (78.885 bhp), and finally, cars with 4 seats have the lowest average maximum power of 36.908 bhp. This suggests that larger vehicles with more seats tend to have higher power output.

Year 2011 to 2020:

Table: Cars data from decade 2011 to 2020

No of seats	Total cars	Average of engine(CC)	Average of mileage(km/l)	Average of max_power(bhp)
2	2	2523	0	70
4	43	1259.28	21.04	97.36
5	5164	1327.64	20.9	90.89
6	62	1635.56	20.01	88.94
7	1004	2043.56	16.16	113.12
8	164	2068.21	14.29	97.48
9	67	2445.63	14.89	80.29
10	11	2352.91	13.08	75.28

Here are some valuable insights from the updated data for the years 2011 to 2020:

Seat Distribution: Cars with 5 seats continue to be the most common choice, accounting for 79.24% of the total cars. This is followed by cars with 7 seats (15.41%), 8 seats (2.52%), 9 seats (1.03%), 6 seats (0.95%), 4 seats (0.66%), and the least common are cars with 2 seats (0.03%). This suggests that the majority of cars during this period were designed for small to mid-sized families.

Engine Size and Seat Count: The average engine size, measured in CC, shows some variation based on the number of seats. Cars with 9 seats have the largest average engine size of 2445.63 CC, followed by cars with 10 seats (2352.91 CC), 8 seats (2068.21 CC), 7 seats (2043.56 CC), 6 seats (1635.56 CC), 2 seats (2523 CC), 5 seats (1327.64 CC), and finally, cars with 4 seats have the smallest average engine size of 1259.28 CC. This suggests that larger vehicles with more seats tend to have larger engines.

Mileage and Seat Count: The average mileage, measured in km/l (kilometres per litre), also shows some variation across different seat counts. Cars with 7 seats have the highest average mileage of 16.16 km/l, followed by cars with 6 seats (20.01 km/l), 5 seats (20.9 km/l), 4 seats (21.04 km/l), 8 seats (14.29 km/l), 9 seats (14.89 km/l), 10 seats (13.08 km/l), and finally, cars with 2 seats have the lowest average mileage of 0 km/l. It's worth noting that the average mileage for cars with 2 seats is likely influenced by outliers or special-purpose vehicles with low or no mileage figures available.

Power and Seat Count: The average maximum power output, measured in bhp (brake horsepower), also shows variation based on the number of seats. Cars with 7 seats have the highest average maximum power of 113.12 bhp, followed by cars with 4 seats (97.36 bhp), 8 seats (97.48 bhp), 6 seats (88.94 bhp), 9 seats (80.29 bhp), 10 seats (75.28 bhp), 5 seats (90.89 bhp), and finally, cars with 2 seats have an average maximum power of 70 bhp.

Comparing the data from two decades (2001-2010 and 2011-2020) can provide additional insights into the changes and trends in vehicle seat count, engine size, mileage, and power over time.

Here are some valuable insights:

1. Seat Distribution: In both decades, cars with 5 seats were the most common choice, indicating their popularity among consumers. However, there is a slight decrease in the percentage of cars with 5 seats in the second decade (78.72% in the first decade vs. 79.24% in the second decade). On the other hand, cars with 7 seats showed an increase in the second decade (9.2% in the first decade vs. 15.41% in the second decade). This suggests a possible trend towards larger family-oriented vehicles in the later decade.

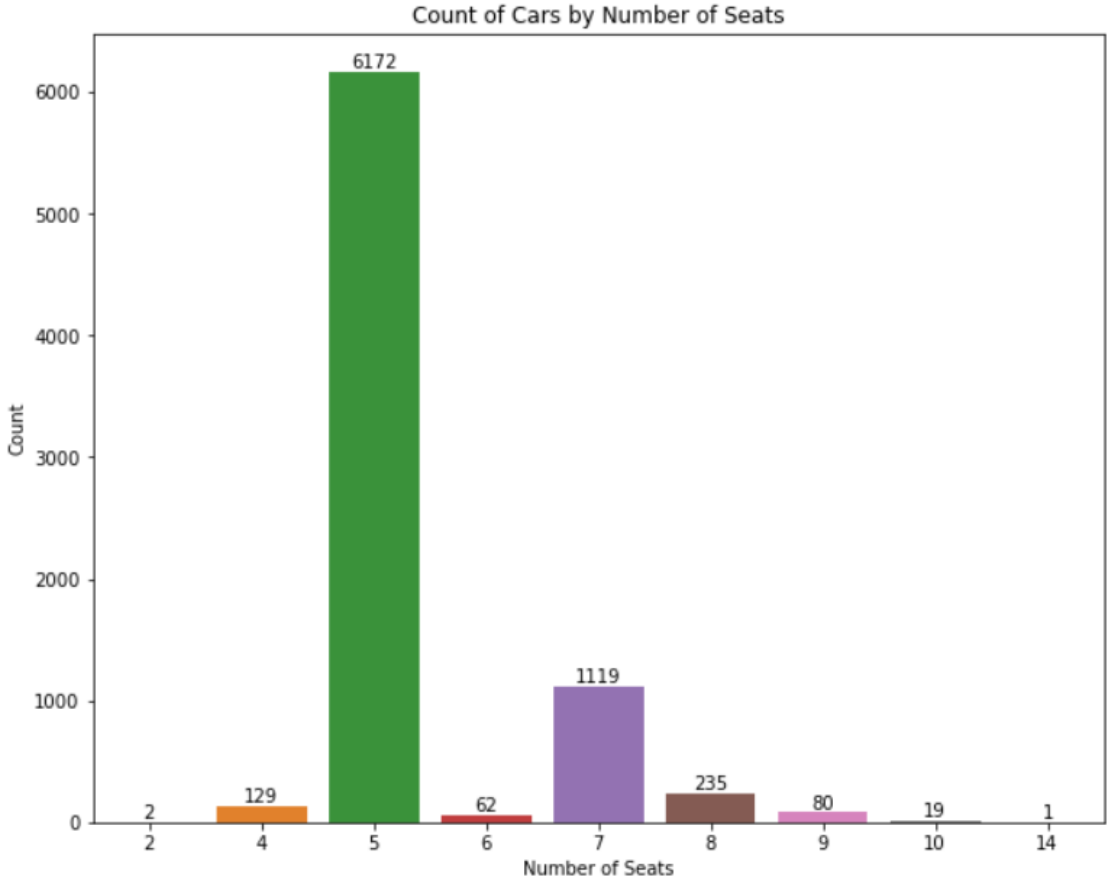


Fig: Overall seats distribution

2. Engine Size: The average engine size, measured in CC, showed slight variations between the two decades. In the first decade, cars with 9 seats had the largest average engine size (2283.923 CC), while in the second decade, cars with 9 seats had the largest average engine size (2445.63 CC). Overall, there seems to be a slight increase in average engine size across different seat counts in the second decade, indicating a possible preference for more powerful vehicles.

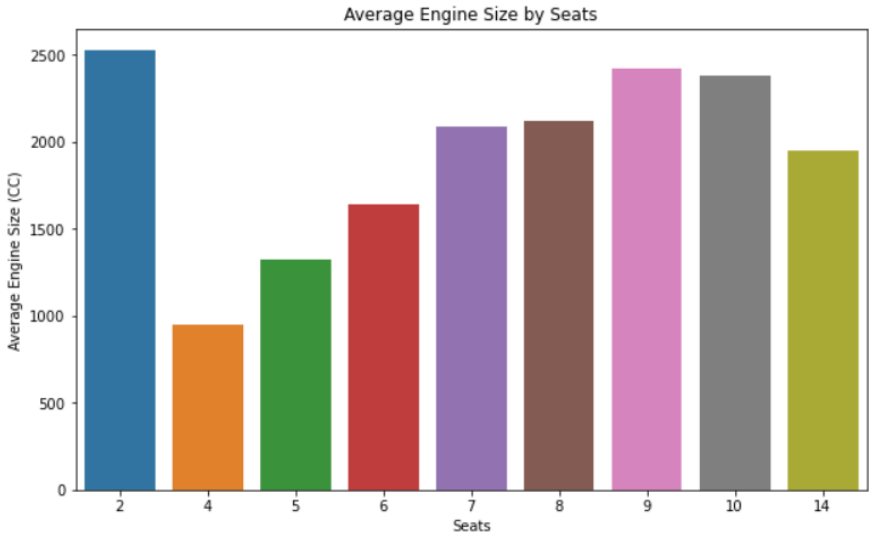


Fig: Average Engine size seat wise

3. Mileage: The average mileage, measured in km/l, exhibited variations between the two decades. In the first decade, cars with 5 seats had the highest average mileage (17.581 km/l), while in the second decade, cars with 7 seats had the highest average mileage (16.16 km/l). The overall trend suggests a slight decrease in average mileage across different seat counts in the second decade, which could be attributed to the increase in vehicle size and power.

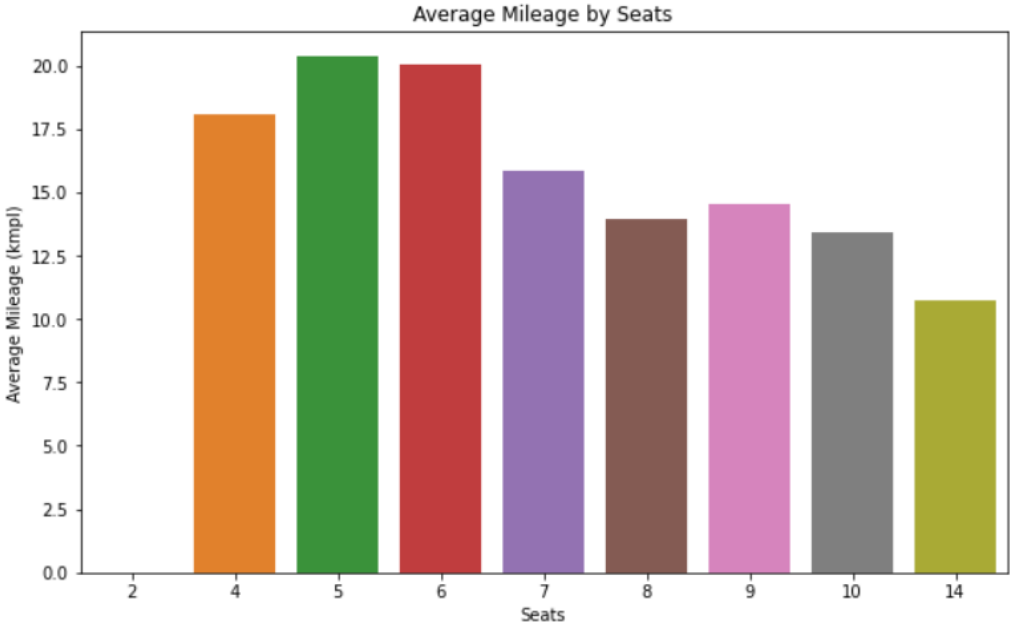


Fig: Average Mileage size seat wise

4. Power: The average maximum power output, measured in bhp, also showed some differences between the two decades. In the first decade, cars with 7 seats had the highest average maximum power (110.995 bhp), while in the second decade, cars with 7 seats again had the highest average maximum power (113.12 bhp). The trend suggests a slight increase in average power output for cars with 7 seats in the second decade, further indicating a preference for more powerful vehicles

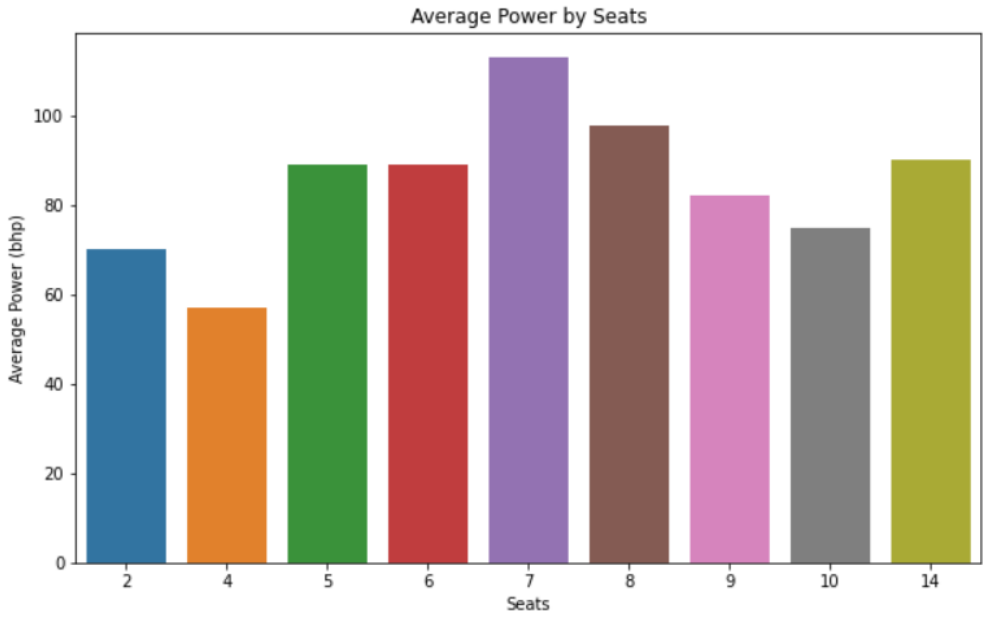


Fig: Average power by seats

These insights highlight the changes in seat distribution, engine size, mileage, and power between the two decades. While the popularity of cars with 5 seats remained consistent, there was an increase in larger vehicles with 7 seats. Additionally, there seems to be a trend towards larger engine sizes and higher power outputs in the later decade, possibly at the expense of fuel efficiency. These insights provide valuable information on the evolving preferences and characteristics of vehicles over time.

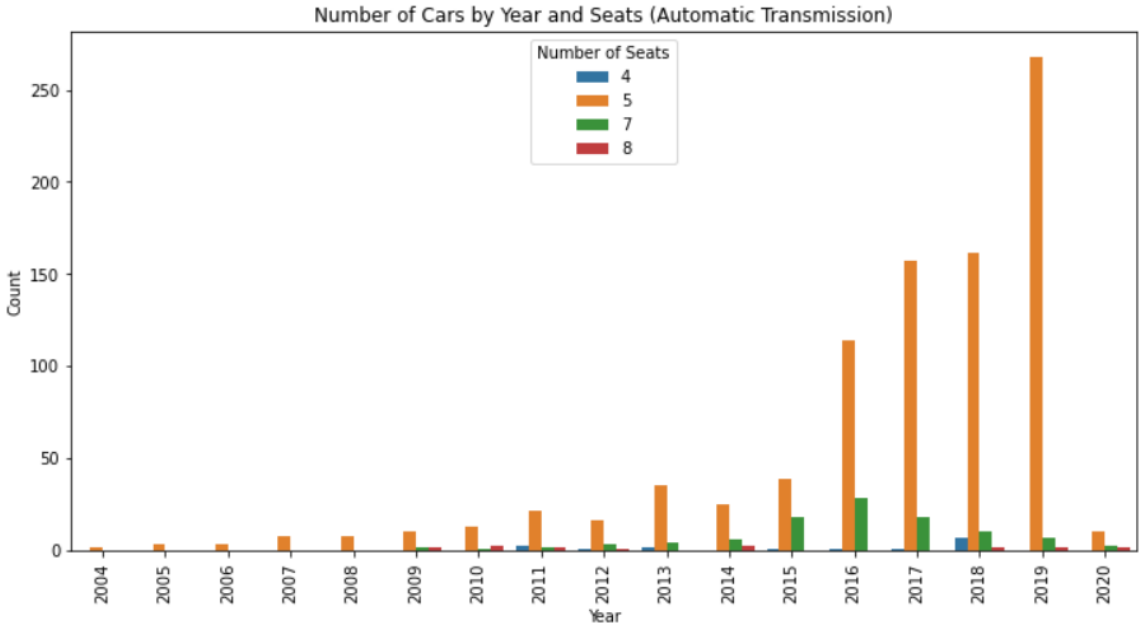


Fig: Number of cars by year and seats for Automatic transmission

From the above figure we can observe there is an increasing trend in buying automatic transmission cars and demand is more for 5 seats cars.

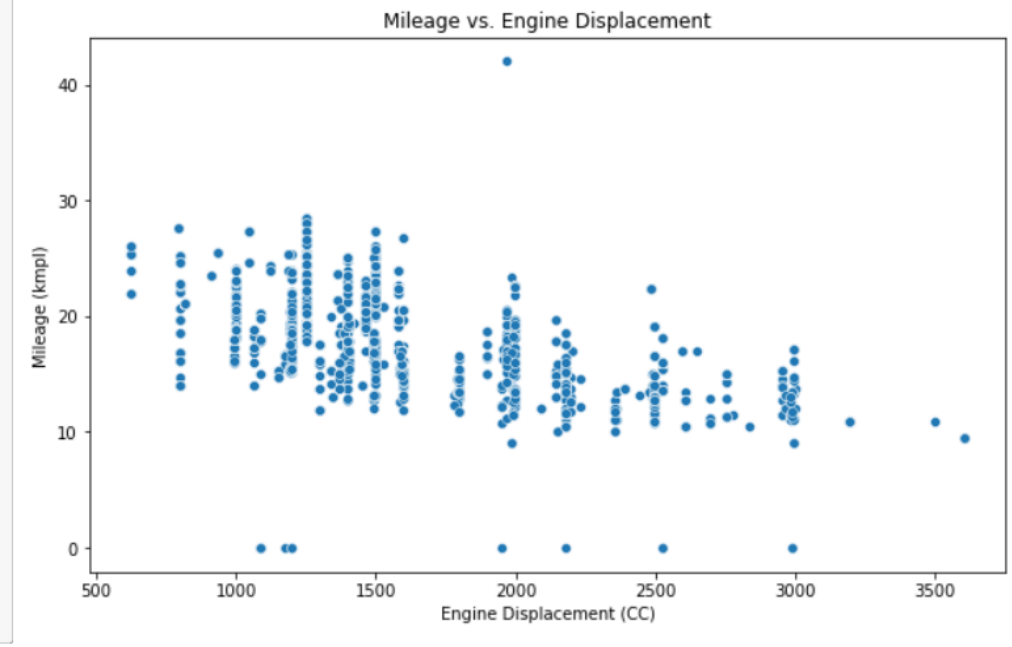


Fig: Scatter plot (Mileage vs Engine capacity)

Scatter plot is plotted for observing trends between mileage and engine capacity. From the above figure we can provide a statement regarding cars that there is a decreasing trend. As engine capacity is increasing there is a slight reduction in mileage.

Some of the data is collected from Indian government sites regarding the EV market in India.

Operational Electric Vehicle (EV) charging station by State/UT:

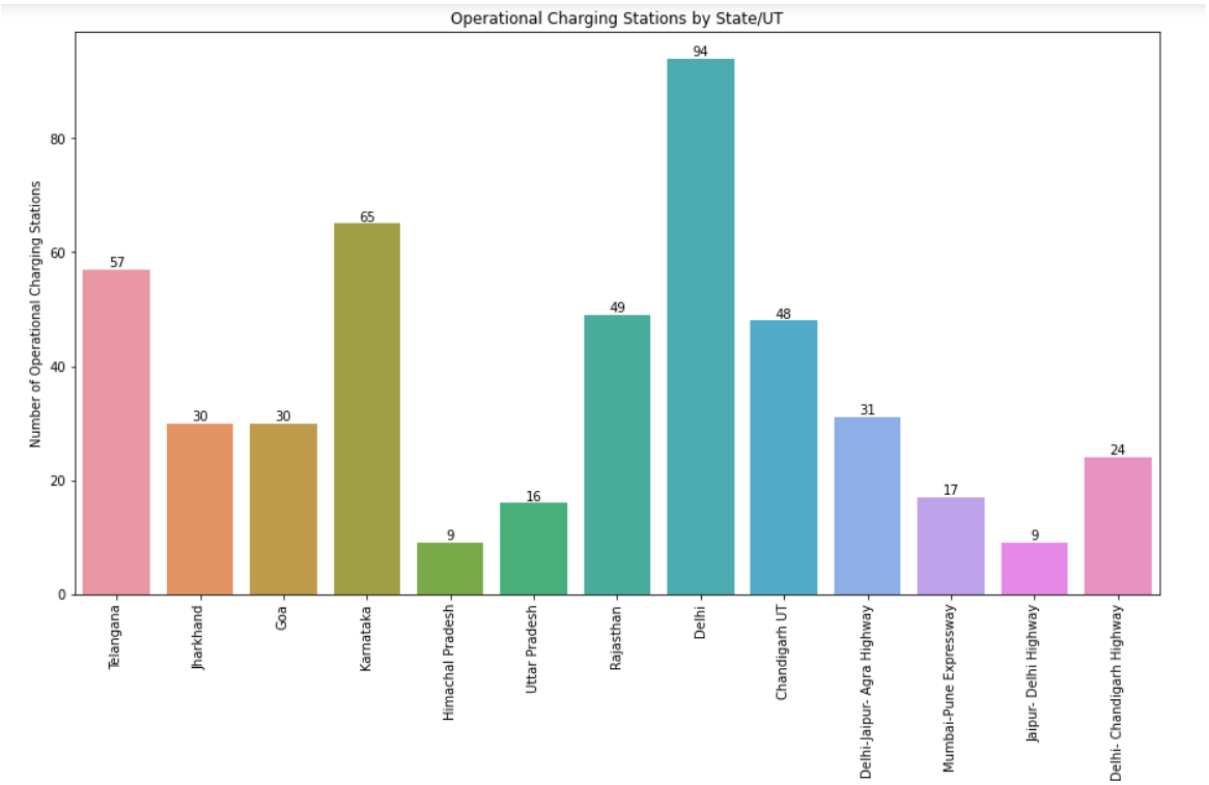


Fig: Operating charging stations by State/UT

Here are some of the insights from the data sets:

- Delhi has the highest number of operational charging stations with 94, followed by Telangana with 57 and Rajasthan with 49.
- The states of Goa, Jharkhand, and Delhi-Jaipur-Agra Highway have an equal number of operational charging stations, each with 30.
- Himachal Pradesh and Jaipur-Delhi Highway have the lowest number of operational charging stations, both with only 9.
- Mumbai-Pune Expressway and Uttar Pradesh have relatively fewer operational charging stations, with 17 and 16 respectively.
- Telangana, Karnataka, and Rajasthan have a relatively higher number of operational charging stations compared to other states.
- Delhi, being the capital city, has a significant number of operational charging stations, which indicates the growing adoption of electric vehicles in the region.
- The number of operational charging stations varies across different states and highways, suggesting differences in infrastructure development and government initiatives for electric vehicle adoption.

Electric Vehicle (EV) chargers sanctioned by State/UT:

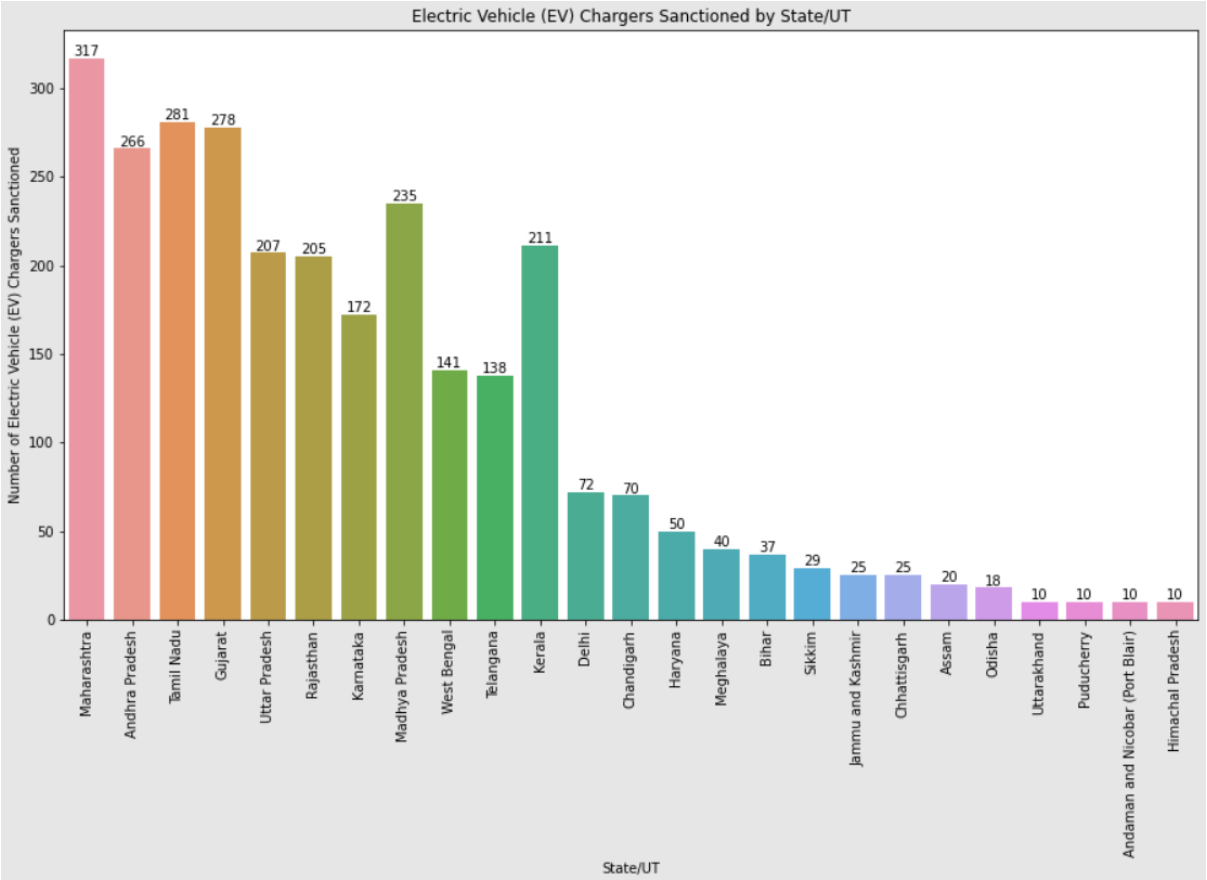


Fig: Electric vehicle (EV) chargers sanctioned by State/UT

Based on the provided data of the number of Electric Vehicle (EV) chargers sanctioned by State/UT, here are some valuable insights:

1. Maharashtra has the highest number of EV chargers sanctioned with 317, followed by Tamil Nadu with 281 and Gujarat with 278.

2. Andhra Pradesh, Madhya Pradesh, and Kerala also have a significant number of EV chargers sanctioned, each with over 200 chargers.

3. States like Jammu and Kashmir, Chhattisgarh, Assam, Odisha, and Bihar have a relatively lower number of sanctioned EV chargers, ranging from 18 to 37.

4. Delhi, despite being the national capital, has a relatively lower number of sanctioned EV chargers with 72, indicating potential for further infrastructure development in the city.

5. The distribution of sanctioned EV chargers varies across states, reflecting differences in government initiatives, policies, and the level of electric vehicle adoption.

6. Maharashtra, Andhra Pradesh, Tamil Nadu, and Gujarat emerge as the leading states in terms of sanctioned EV chargers, indicating a proactive approach towards promoting electric vehicle infrastructure.

These insights provide a general overview of the distribution of sanctioned EV chargers among the mentioned states/UT. Further analysis could involve comparing these numbers with factors such as population, electric vehicle sales, or government incentives to gain a more comprehensive understanding of the electric vehicle charging infrastructure landscape.

Electric Vehicle (EV) sales/invoice by State/UT:

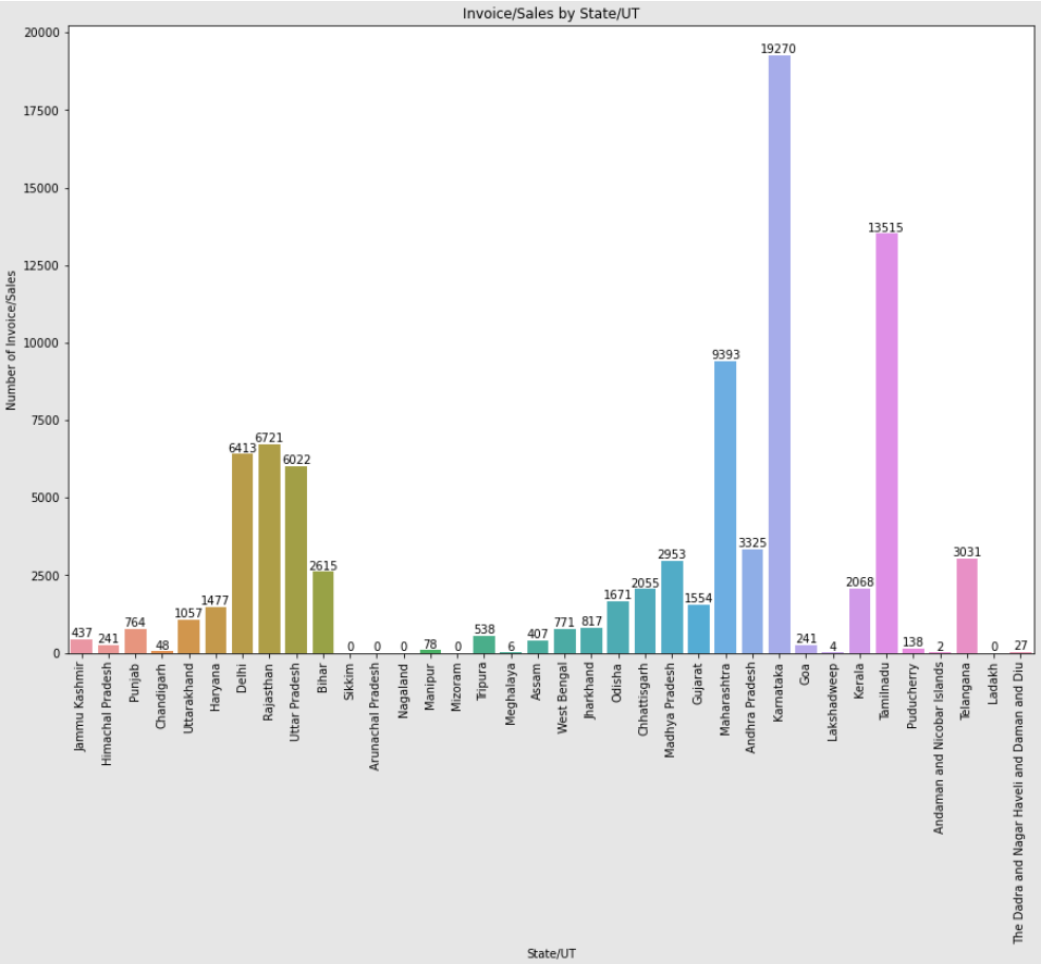


Fig: Electric vehicle (EV) sales/invoice by State/UT

Based on the provided data of the number of invoice/sales by State/UT, here are some valuable insights:

1. Karnataka has the highest number of invoice/sales with 19,270, followed by Tamil Nadu with 13,515 and Maharashtra with 9,393.

2. Uttar Pradesh, Rajasthan, and Delhi also have a significant number of invoice/sales, each with over 6,000.

3. States like Bihar, Madhya Pradesh, and Chhattisgarh have moderate numbers of invoice/sales, ranging from 2,615 to 2,953.

4. Sikkim, Arunachal Pradesh, Nagaland, Mizoram, and Ladakh have no reported invoice/sales in the provided data.

5. Kerala, Odisha, and Telangana have a relatively higher number of invoice/sales, ranging from 1,671 to 3,031.

6. The distribution of invoice/sales varies across different states/UT, indicating differences in economic activity and market demand.

7. Karnataka, Tamil Nadu, and Maharashtra emerge as the leading states in terms of invoice/sales, indicating robust business and commercial activity.

Vikrant Arora :

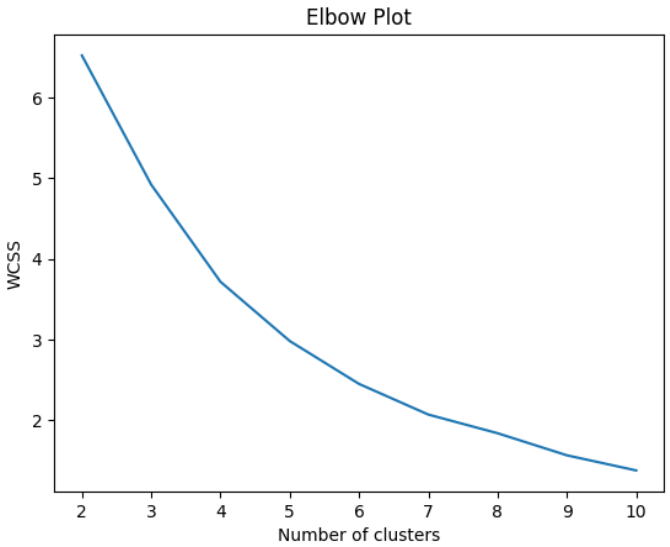
1. Dataset:
https://drive.google.com/file/d/1mshWg1hYSG5ifT3988eDCeGTWnSd8Z2n/view?usp=sharing
The dataset is that of the Indian vehicle market, not differentiated by fuel or electric vehicles.

2. Data Preprocessing:
Libraries used:
- pandas
- sklearn.preprocessing

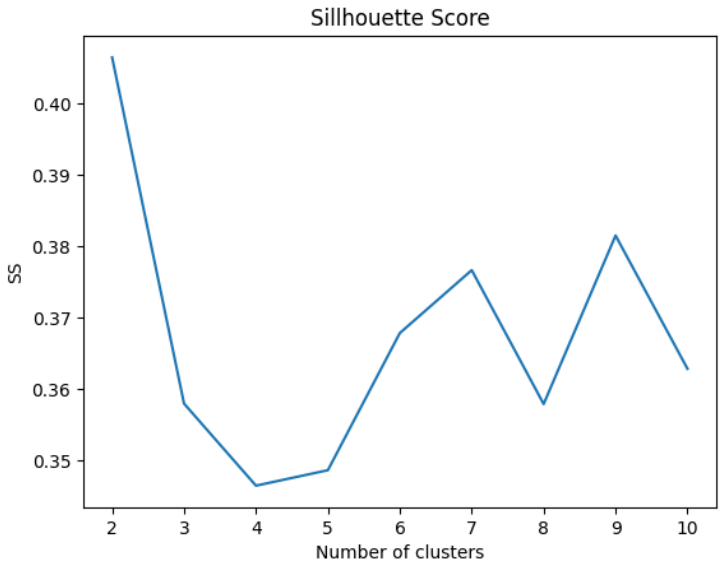
The dataset by itself is mostly ready-to-use; it did not contain any missing values or particularly unnatural outliers. Most of the preprocessing involved selecting columns to use for the clustering process and scaling the individual data points.

3. Segment Extraction
Libraries used:
- pandas
- sklearn.kmeans
- sklearn.sillhouette_score
- matplotlib.pyplot
- seaborn

The primary method of segmentation was K-Means clustering as provided by the scikit_learn API. Segment stability was judged by the use of the elbow method and the silhouette score.



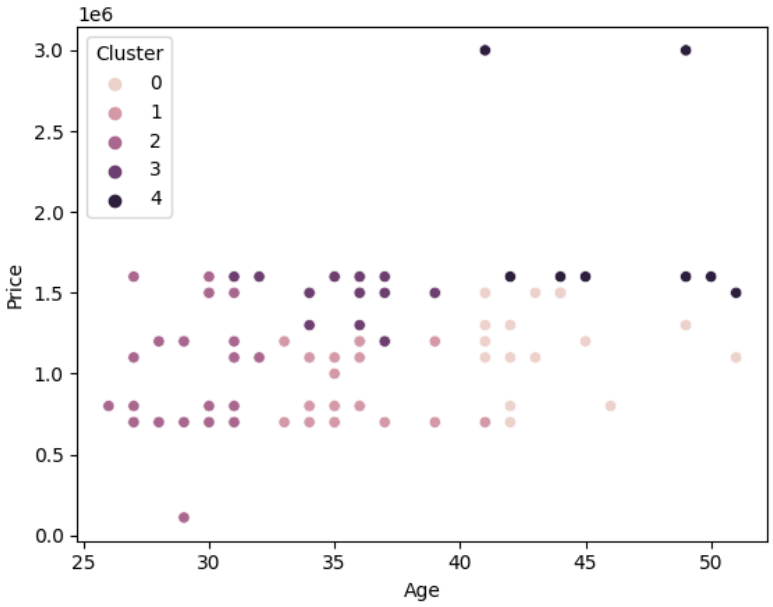
The within cluster sum of squares[WCSS] has an orderly descent with an elbows formed at 4, 5 clusters respectively. As is visible, the difference in slopes is not as steep, hence the silhouette score was also employed to decide on a credible number of clusters. SIlhouette score refused to converge at a single value, alternating between the number of clusters from 5 to 8 with varying magnitudes, also yielding surprisingly low values for these very numbers for certain iterations, but generally agreeing that 5 to 8 is the ideal number.



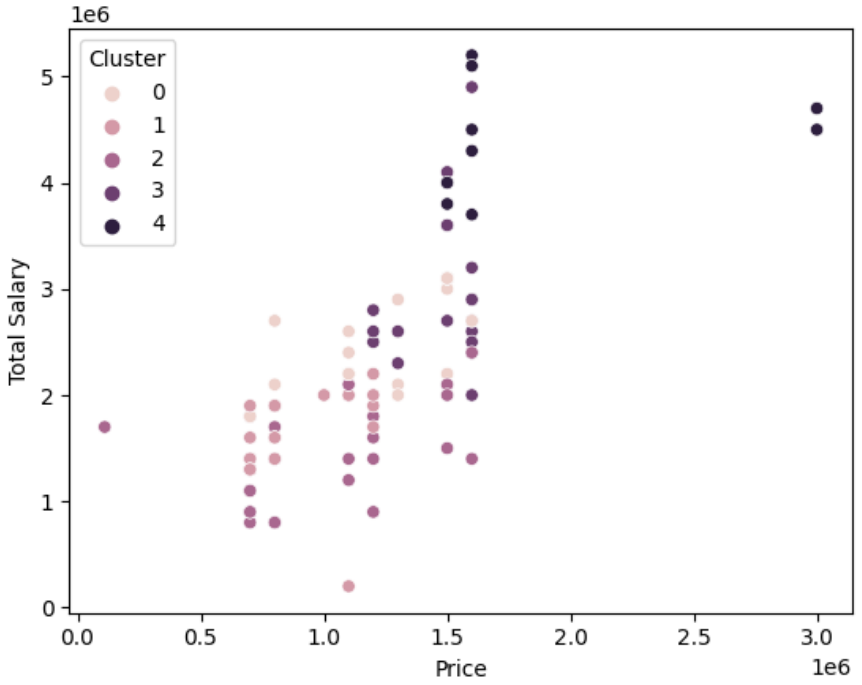
From the following graph of a particular iteration, one might be tempted to use 9, 6 or 7 as k, but since the number of rows was very limited(99), 5 felt like a more appropriate choice.

4. Profiling Potential Segments

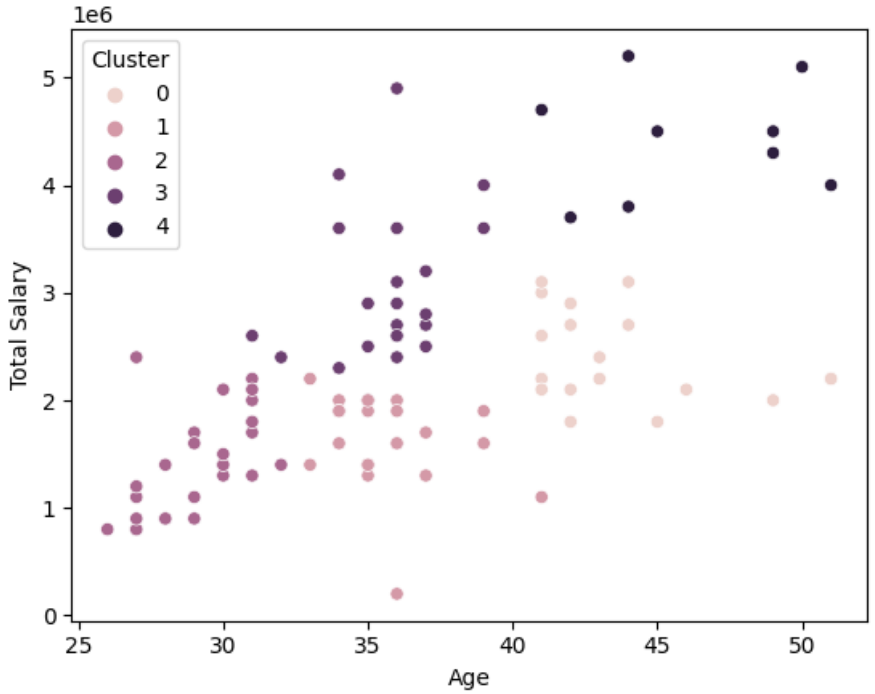
After completion of the clustering process, different quantities were used to make scatter plots, using the cluster label for the hues.



The above diagram depicts neatly formed clusters in the distribution of age and price. The clusters do not map as neatly onto other plots, but act as projections onto other quantities when one of the axes is retained, functioning like an internal join.

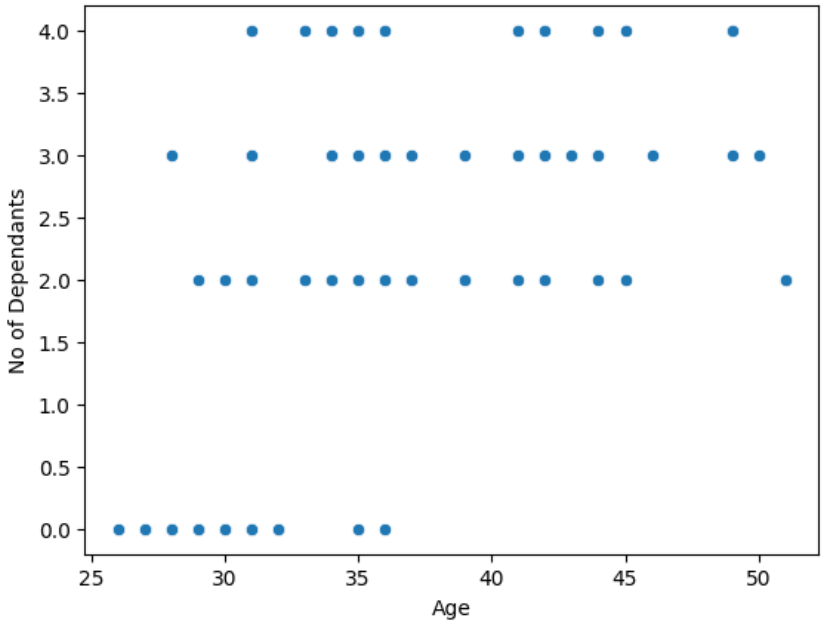


One major albeit obvious observation is that the most preferred price category is mid range, between 10 and 17 lakh rupees, regardless of the total salary, barring a few outliers.
It is a general observation that people of a younger strata are more receptive to electric vehicles and more likely to buy them than their older counterparts(not proven in this current study, but is widely agreed upon inside the market). Hence a more favourable segment age-wise would be on the younger side, with a medium to high total salary.



5. Selecting target segment

Segment number 3 of the above clustering report fulfils these requirements, although it does have an older demographic by comparison, between 33 and 40, but they have a higher income than the younger segments, making them more open to purchasing higher end vehicles.



This described age group on average has a sizable average number of dependents, and thus would be more open to a bigger vehicle for transport purposes.

This makes it quite evident that it would be the most profitable segment to market to.

6. Customising market mix

To market to the selected market segment, larger vehicles would be more favourable, owing to the higher number of dependants. A moderately steep price range of around 13 to 20 lakhs would not be a major problem as this segment is on par with the purchasing power of older segments, while still being more open to electric vehicles.

Kamlesh Prajapat:

Data set link https://www.kaggle.com/datasets/surajsunilbhosale/electric-vehical-charging-station

Libraries Used

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Sklearn

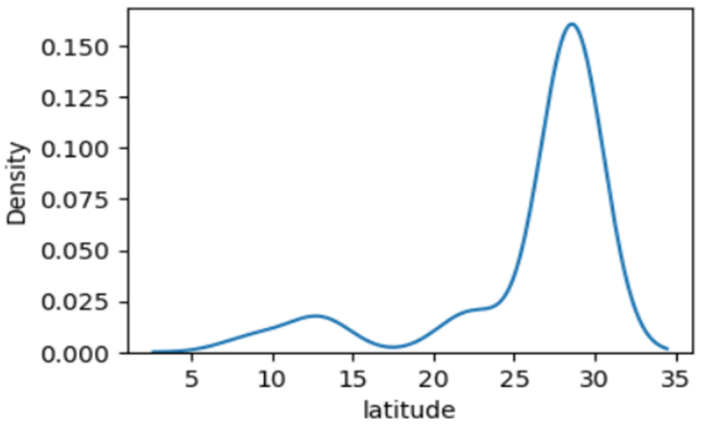
Segment Extraction

SVM models are versatile tools for classifying vehicles into the EV market or non-EV market, aiding in market share analysis and assessing adoption rates. They can also be used for regression analysis to predict sales volumes and demand for electric vehicles, supporting strategic decision-making. SVM models provide valuable insights into the EV market's current state, penetration levels, and growth potential. These applications assist stakeholders in understanding market dynamics and identifying opportunities for the development and advancement of the EV industry.

Spatial Patterns: The utilisation of latitude and longitude as predictors in the SVM model suggests an exploration of spatial patterns in the electric vehicle market.

Regional Differences: The accuracy score of 63.41% indicates that the SVM model has some predictive capability for classifying the regions based on latitude and longitude.

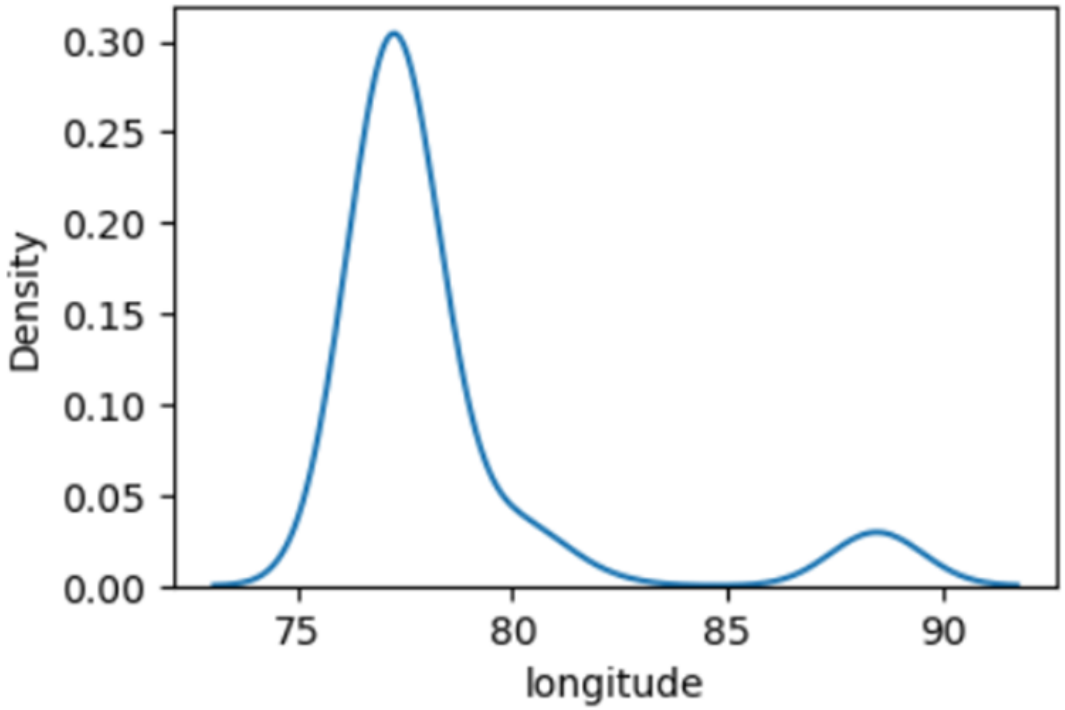
Latitude: Latitude is a geographic coordinate that specifies the north-south position of a location on Earth. It is measured in degrees, with values ranging from -90° at the South Pole to +90° at the North Pole. The equator is defined as 0° latitude.



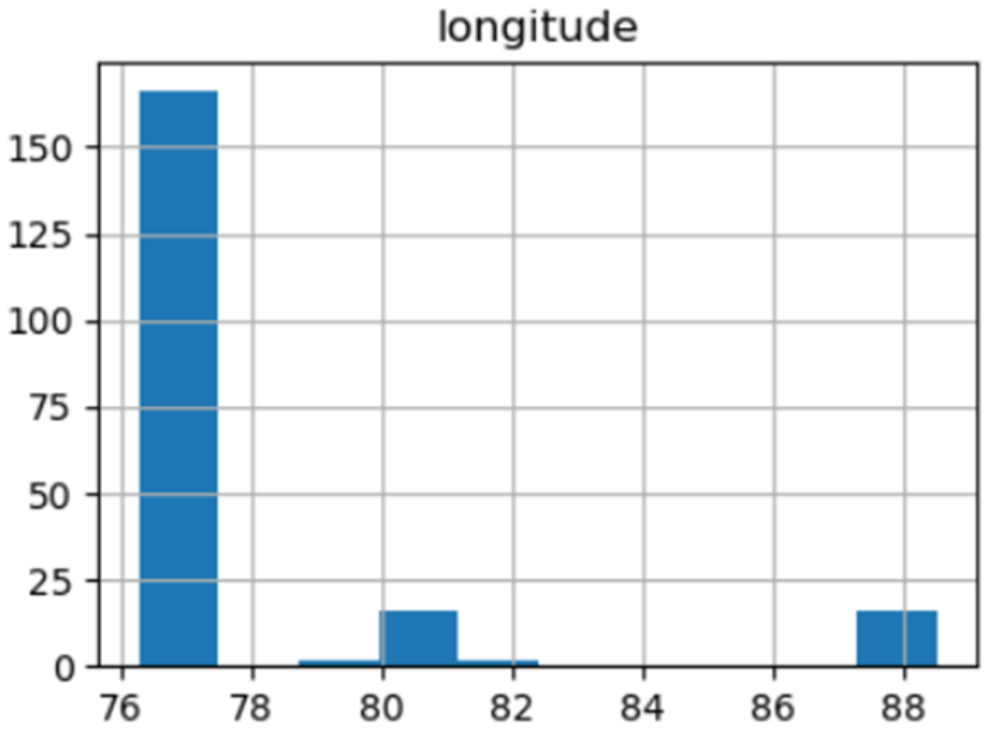
By observing the provided data, it can be noted that a significant portion of the latitude values falls within the range of 24 to 34. This range encompasses the majority of latitude measurements recorded in the dataset. The prevalence of latitude values within this range suggests a concentration of data points in geographical regions situated between 24 and 34

degrees latitude. This finding indicates a potential focus or emphasis on specific geographic areas or regions associated with the dataset's context, likely related to the electric vehicle market.

Longitude: Longitude is a geographic coordinate that denotes the east-west position of a location on Earth. It is also measured in degrees, with values ranging from -180° to +180°. The prime meridian, which passes through Greenwich, London, is defined as 0° longitude.



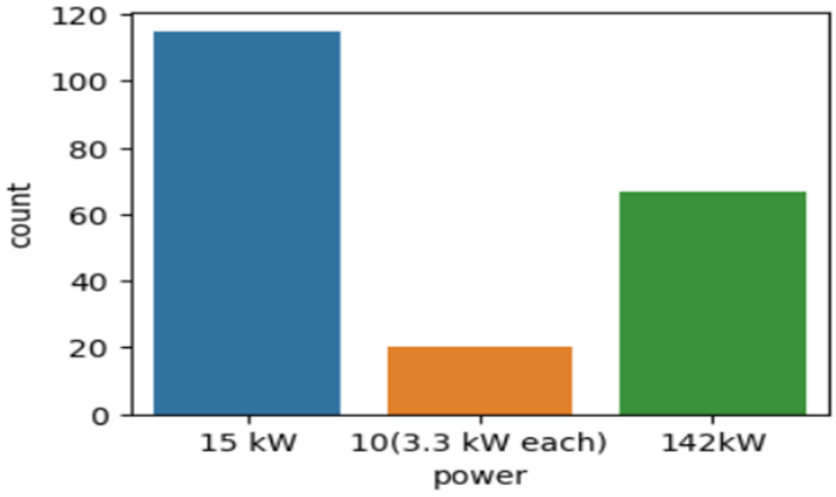
Based on the data set, it is evident that a substantial portion of the longitude values lies within the range of 75 to 80. This range encapsulates the majority of longitude measurements recorded in the dataset. The prevalence of longitude values within this range indicates a concentration of data points in geographic regions located between 75 and 80 degrees longitude. This finding suggests a focus or emphasis on specific areas or regions associated with the dataset's context, potentially related to the electric vehicle market. These regions with longitude values between 75 to 80 may exhibit particular characteristics, such as high electric vehicle adoption, significant infrastructure development, or other relevant factors influencing the market dynamics.



Based on the analysis of the provided diagram, we can observe a clear pattern indicating the prevalent power rating among electric vehicles in the dataset. The diagram reveals that a significant majority of electric vehicles utilise a power rating of 15 KW. This suggests that vehicles with a 15 KW power battery are more commonly adopted or favoured within the analysed dataset.

In contrast, the diagram also highlights that vehicles equipped with a 10 KW power battery exhibit lower usage or adoption rates. This indicates that electric vehicles with a 10 KW power rating are less commonly chosen or have lower market penetration compared to their 15 KW counterparts.

```
15 kW               115
142kW                67
10(3.3 kW each)      20
Name: power, dtype: int64
```



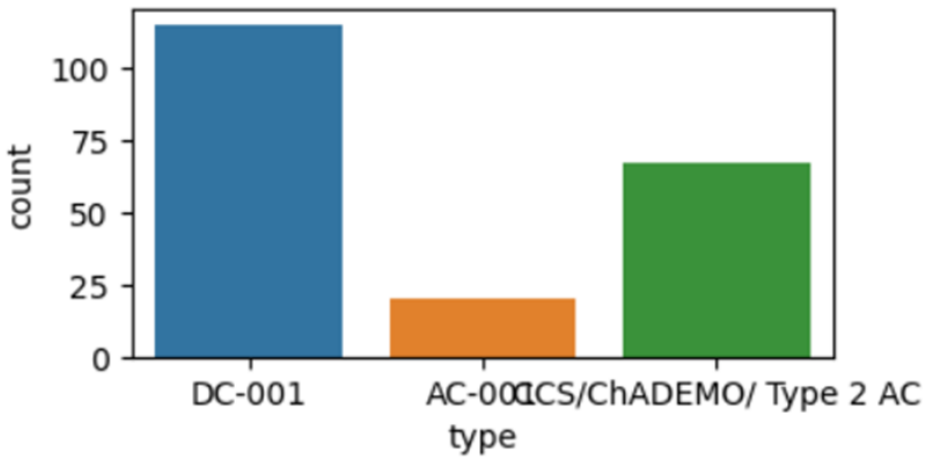
The dominance of the 15 KW power rating among the electric vehicles in the dataset implies a stronger preference or demand for vehicles with higher power capacity. This preference could be attributed to various factors, such as enhanced performance, extended range, or faster charging capabilities associated with the 15 KW power rating. It suggests that electric vehicle users or buyers within the analysed dataset tend to opt for vehicles that offer higher power capacity.

This insight is valuable for understanding the market dynamics and consumer preferences within the electric vehicle industry. It provides useful information for manufacturers, policymakers, and stakeholders involved in the design, development, and marketing of electric vehicles. Manufacturers can prioritise the production of vehicles with 15 KW power batteries to meet the prevalent demand and capitalise on the market trend. Policymakers can consider promoting incentives or initiatives that encourage the adoption of electric vehicles with higher power capacity to align with consumer preferences.

From the diagram provided, a clear observation can be made that the majority of the electric vehicles in the dataset are classified as type "DC-001". This indicates that the "DC-001" type has a higher prevalence among the analysed volume of electric vehicles.

The dominance of the "DC-001" type suggests that it is the most commonly represented or chosen electric vehicle type within the dataset. This finding can have several implications, such as:

```
DC-001                    115
CCS/ChADEMO/ Type 2 AC     67
AC-001                     20
Name: type, dtype: int64
```



Manufacturing Focus: The higher prevalence of the "DC-001" type suggests that manufacturers may be placing a significant emphasis on producing and promoting this specific electric vehicle model. It could indicate that the "DC-001" type is a popular choice among consumers or that manufacturers have invested more resources in its production.

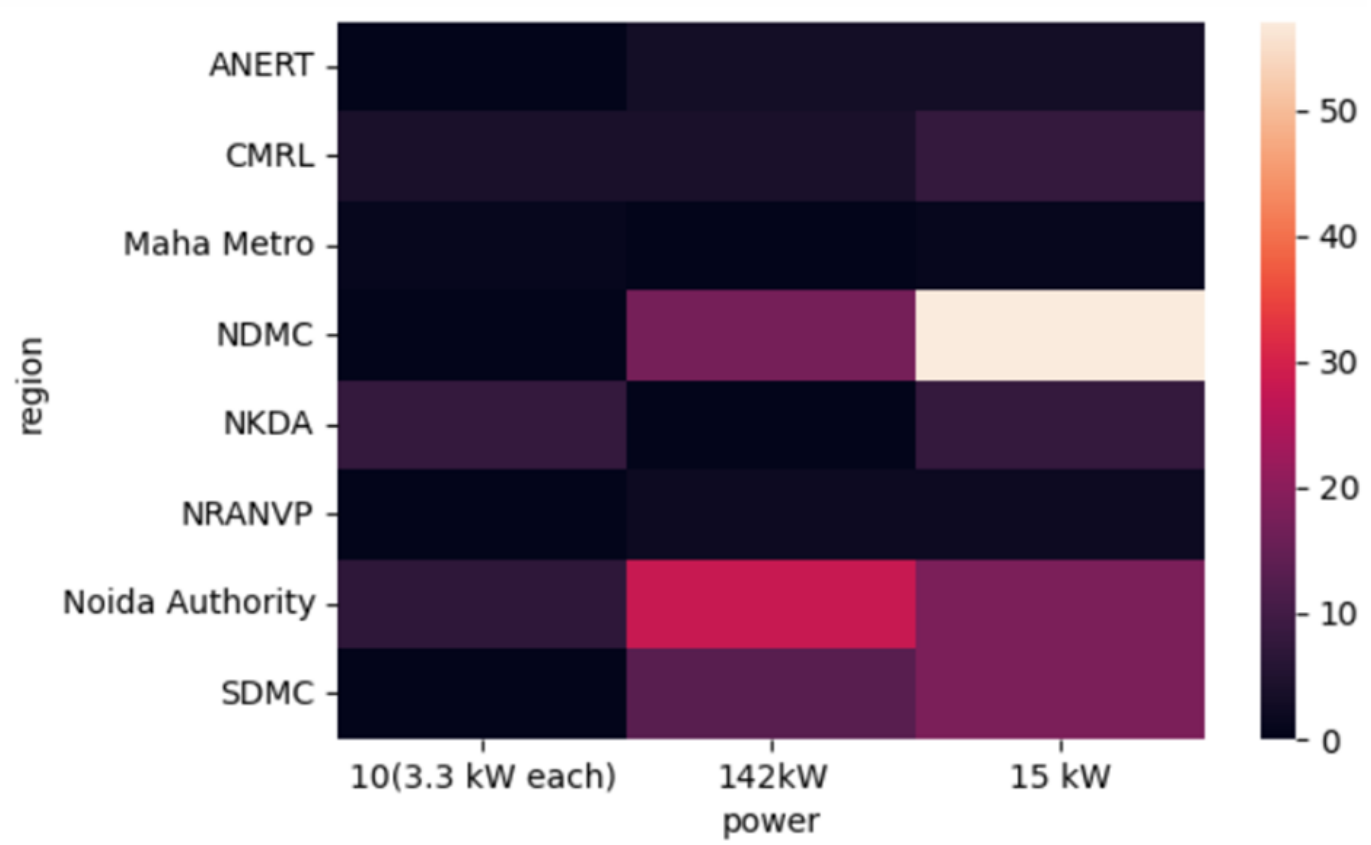
Consumer Demand: The popularity of the "DC-001" type indicates a higher demand or preference among electric vehicle buyers within the dataset. This could be due to factors such as pricing, features, performance, range, or overall customer satisfaction associated with the "DC-001" type.

Market Positioning: The dominance of the "DC-001" type may signify a strategic positioning by the manufacturer in the electric vehicle market. They might have identified specific market segments or target audiences that align well with the features and attributes of the "DC-001" type, leading to its higher representation.

Industry Trends: The prevalence of the "DC-001" type could reflect broader industry trends and advancements. It may indicate technological innovations, design preferences, or regulatory compliance that make the "DC-001" type a prominent choice in the evolving electric vehicle market.

Understanding the popularity of the "DC-001" type within the dataset is valuable for manufacturers, consumers, and industry analysts. Manufacturers can leverage this information to continue improving and promoting the "DC-001" type or gain insights into what attributes make it successful. Consumers can use this knowledge to consider the reputation, performance, and reliability of the "DC-001" type when making purchasing decisions. Additionally, industry analysts can examine the market dynamics and competitive landscape surrounding the "DC-001" type to identify opportunities or assess market trends.

power region	10(3.3 kW each)	142kW	15 kW
ANERT	0	3	3
CMRL	4	4	8
Maha Metro	1	0	1
NDMC	0	17	57
NKDA	8	0	8
NRANVP	0	2	2
Noida Authority	7	28	18
SDMC	0	13	18



By examining the region-wise distribution of power among electric vehicles, several insights can be extracted from the dataset. Firstly, it can be observed that a significant number of electric vehicles with a power rating of 15 kW are concentrated in the NDMC (New Delhi Municipal Council) region. This indicates a higher prevalence and adoption of 15 kW electric vehicles in the NDMC area. The popularity of this power rating in NDMC suggests a favourable environment or incentives for electric vehicles with 15 kW power capacity.

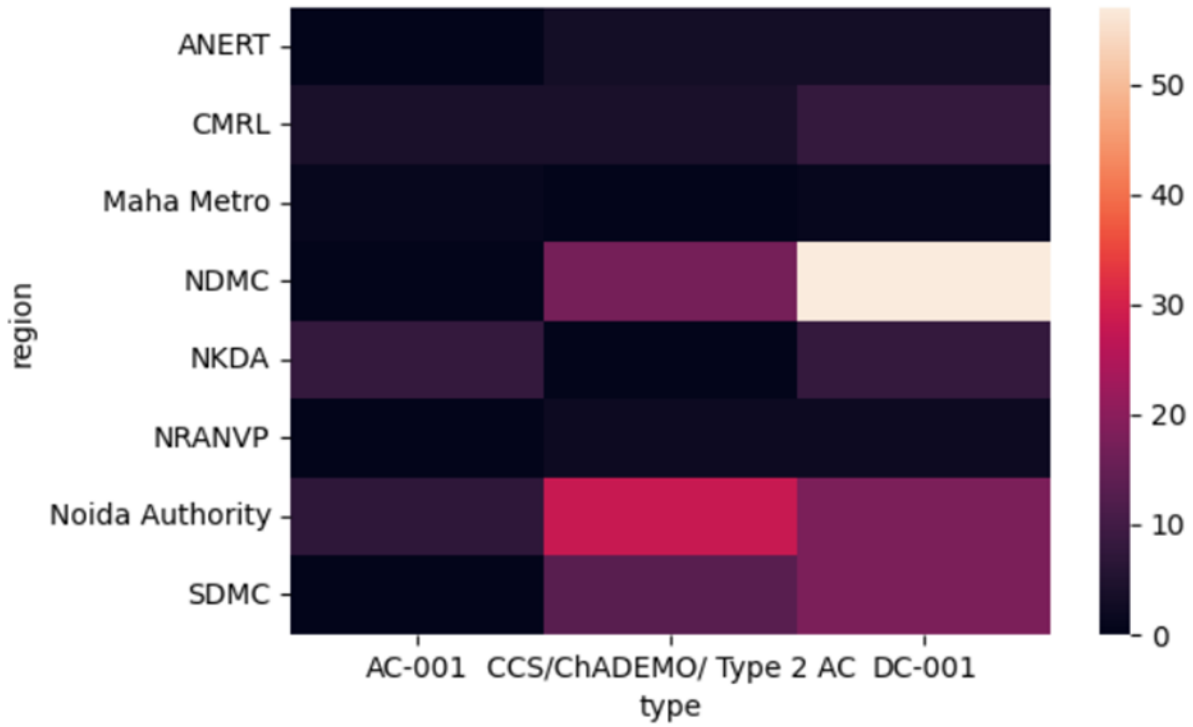
Furthermore, it is noteworthy that NDMC is also the region with the highest number of electric vehicles overall. This implies that NDMC demonstrates a strong inclination towards electric vehicle adoption across various power ratings, indicating a favourable environment for electric mobility.

Another significant observation is that electric vehicles with a power rating of 142 kW are primarily found in the Noida Authority region. This indicates a higher concentration and preference for electric vehicles with 142 kW power capacity in the Noida Authority area. The specific power rating of 142 kW may align with the market demands, preferences, or regulations in the Noida Authority region, leading to its higher representation.

These insights highlight regional variations in power distribution and electric vehicle adoption. The concentration of 15 kW electric vehicles in NDMC suggests the significance of this power rating in that region, while the prevalence of 142 kW electric vehicles in the Noida Authority area points towards specific preferences or factors driving the adoption of electric vehicles with that power rating.

Understanding the region-wise distribution of power in conjunction with electric vehicle adoption is valuable for multiple stakeholders. Manufacturers can focus their production and marketing efforts on power ratings that are popular in specific regions. Policymakers and urban planners can identify regions with higher electric vehicle adoption and tailor infrastructure development and incentives accordingly. Consumers can consider the power ratings prevalent in their region when making purchasing decisions.

type region	AC-001	CCS/ChADEMO/	Type 2 AC	DC-001
ANERT	0		3	3
CMRL	4		4	8
Maha Metro	1		0	1
NDMC	0		17	57
NKDA	8		0	8
NRANVP	0		2	2
Noida Authority	7		28	18
SDMC	0		13	18



Examining the region-wise distribution of types of electric vehicles reveals valuable insights about their prevalence in different areas. It can be observed that a significant number of electric vehicles belonging to the "DC-001" type are concentrated in the NDMC (New Delhi Municipal Council) region. This indicates a higher adoption and prevalence of "DC-001" type electric

vehicles in NDMC. The popularity of this specific type in NDMC suggests a favourable environment or incentives for "DC-001" type electric vehicles, leading to their higher representation.

Furthermore, the analysis also reveals that NDMC has the highest number of electric vehicles overall. This implies that NDMC demonstrates a strong inclination towards electric vehicle adoption, irrespective of the type. The high representation of electric vehicles in NDMC indicates a supportive ecosystem, infrastructure, or policies encouraging electric mobility.

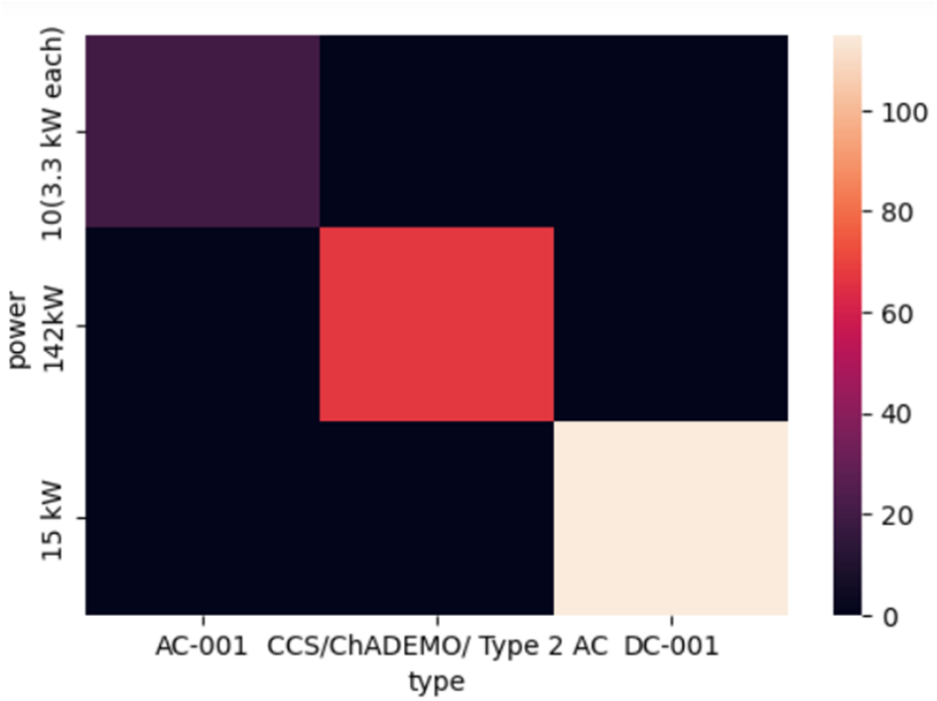
On the other hand, the data shows that most of the electric vehicles of type "2 AC" are found in the Noida Authority region. This indicates a concentration and preference for electric vehicles of type "2 AC" in the Noida Authority area. The specific type "2 AC" may align with the market demands, infrastructure availability, or other factors specific to the Noida Authority region, leading to its higher representation.

These insights highlight regional variations in the distribution of electric vehicle types and adoption. The prevalence of "DC-001" type electric vehicles in NDMC and "2 AC" type electric vehicles in the Noida Authority region suggests the significance of specific types in their respective areas.

Understanding the region-wise distribution of electric vehicle types in relation to their prevalence can guide various stakeholders. Manufacturers can align their production and marketing strategies to cater to the demand for specific types in different regions. Policymakers and urban planners can identify areas with higher adoption of specific types and develop targeted initiatives or infrastructure to support them. Consumers can consider the prevalent types in their region when making purchasing decisions or assessing charging infrastructure compatibility.

type power	AC-001	CCS/ChADEMO/ Type 2 AC	DC-001
10(3.3 kW each)	20	0	0
142kW	0	67	0
15 kW	0	0	115

Based on the previously mentioned insights, a notable observation is that the majority of electric vehicles belonging to the "DC-001" type are associated with a power rating of 15 kW. This finding suggests a strong correlation between the "DC-001" type and the specific power capacity of 15 kW.



The prevalence of 15 kW power capacity among "DC-001" electric vehicles can have several implications:

Manufacturer Emphasis: The concentration of "DC-001" electric vehicles with a power rating of 15 kW implies that manufacturers have intentionally designed and produced this specific type to align with the demand for vehicles of that power capacity. It suggests a deliberate choice by manufacturers to offer a "DC-001" model primarily with 15 kW power, which likely corresponds to the performance capabilities, range, or target market segment for this type.

Market Preference: The significant representation of 15 kW power capacity among "DC-001" electric vehicles indicates a strong preference for this power rating among consumers or within the market segment targeted by the "DC-001" type. It suggests that customers seeking "DC-001" electric vehicles are more inclined to opt for those with 15 kW power capacity, potentially due to factors such as desired range, charging time, or performance associated with this power rating.

Synergy of Type and Power: The alignment between the "DC-001" type and the 15 kW power rating suggests a harmonious combination that meets the specific needs or requirements of the market segment targeted by this type. The 15 kW power capacity may complement the features, intended usage, or market positioning of the "DC-001" type, contributing to its success and prevalence.

Understanding the strong association between the "DC-001" type and the 15 kW power rating provides valuable insights for manufacturers, consumers, and industry stakeholders. Manufacturers can further enhance the "DC-001" type's offerings, focusing on optimising features related to the 15 kW power capacity. Consumers can consider the benefits and limitations associated with the 15 kW power rating when evaluating "DC-001" electric vehicles. Additionally, industry stakeholders can analyse this association to gain a deeper understanding

of market dynamics, preferences, and potential areas for future improvements or product differentiation.

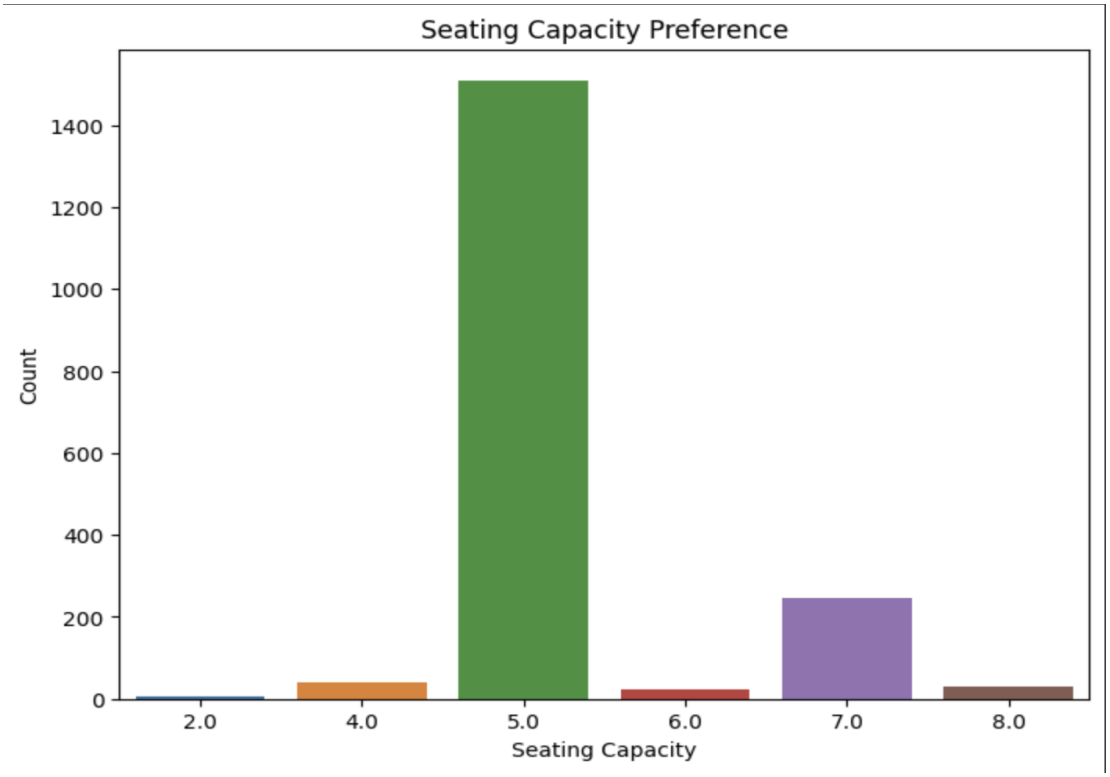
Ruturaj Patil :

Libraries Used:

- Numpy
- Pandas
- Scikit Learn
- Seaborn

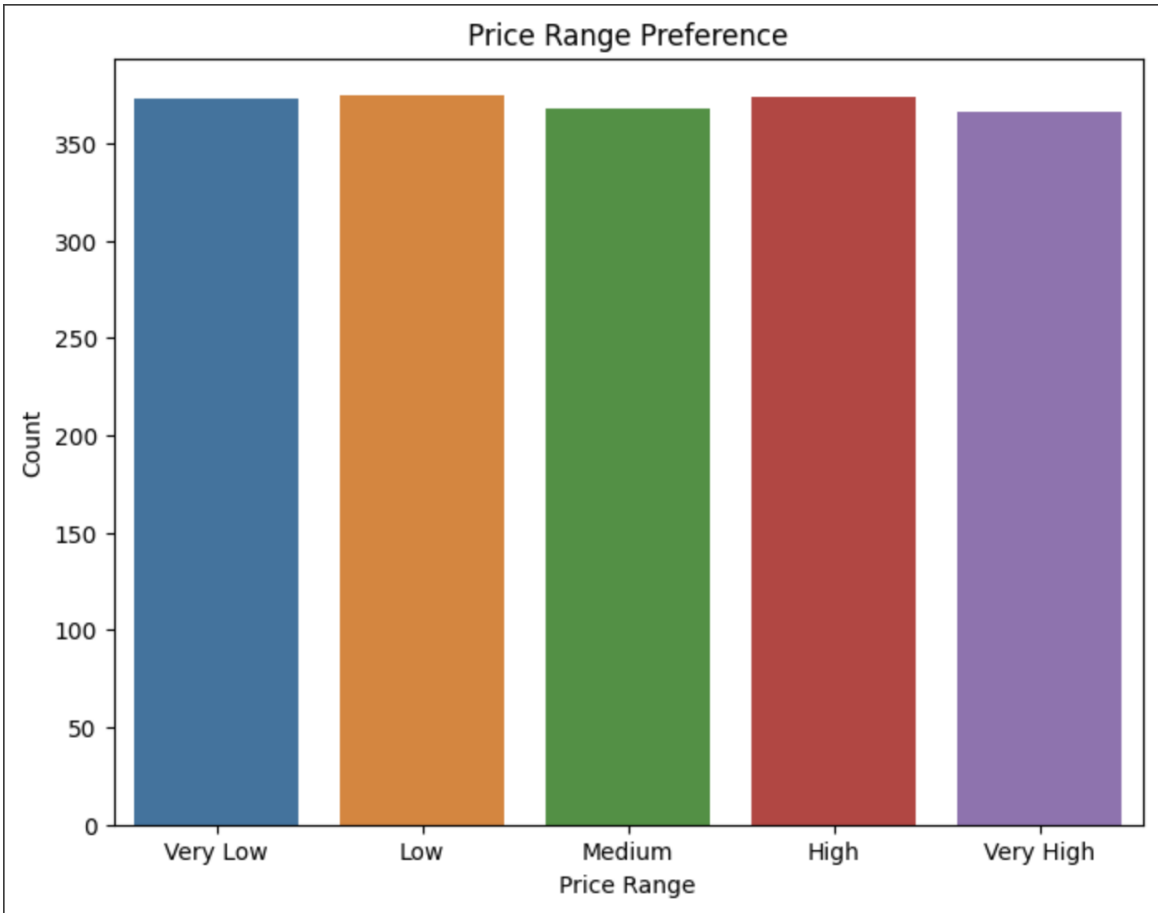
Exploratory Data Analysis:

The dataset comprised comprehensive information on car sales, encompassing key factors such as engine power, torque, length, width, height, fuel tank capacity, seating capacity, and price. In order to conduct an effective market segmentation analysis for the Electric Vehicle (EV) market, certain columns including seller type, fuel type, owner, and kilometres driven were deemed irrelevant and thus excluded from consideration. To ensure data consistency, rows containing missing information pertaining to engine specifications and dimensions were removed from the dataset. This meticulous data curation process aimed to uphold the integrity and reliability of the subsequent market segmentation analysis for the EV market.



To ascertain the prevailing customer preferences in the market regarding car seating capacity, a thorough analysis was conducted. Specifically, a plot was generated comparing the count of car models against their respective seating capacities. The resulting plot provided valuable insights, distinctly indicating that the majority of customers exhibit a strong inclination towards 5-seater cars. This finding underscores the significance of accommodating the demand for vehicles with a

seating capacity of five, thereby facilitating targeted market segmentation strategies for the Electric Vehicle (EV) market.

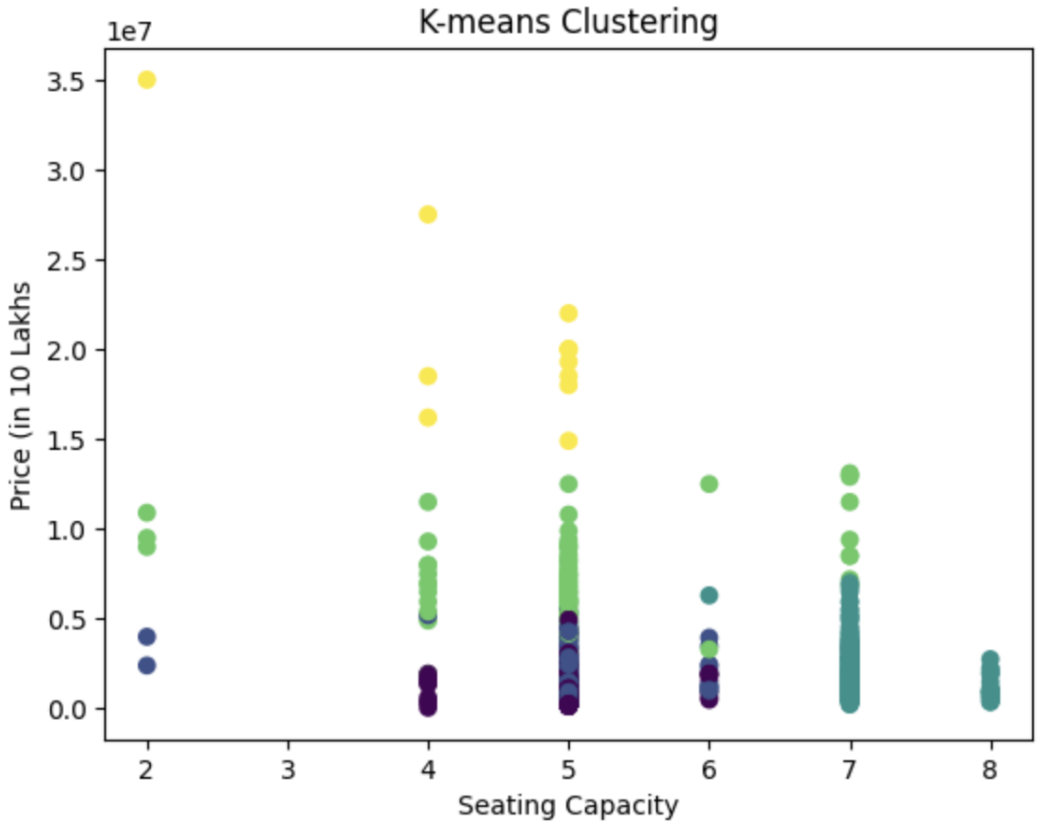


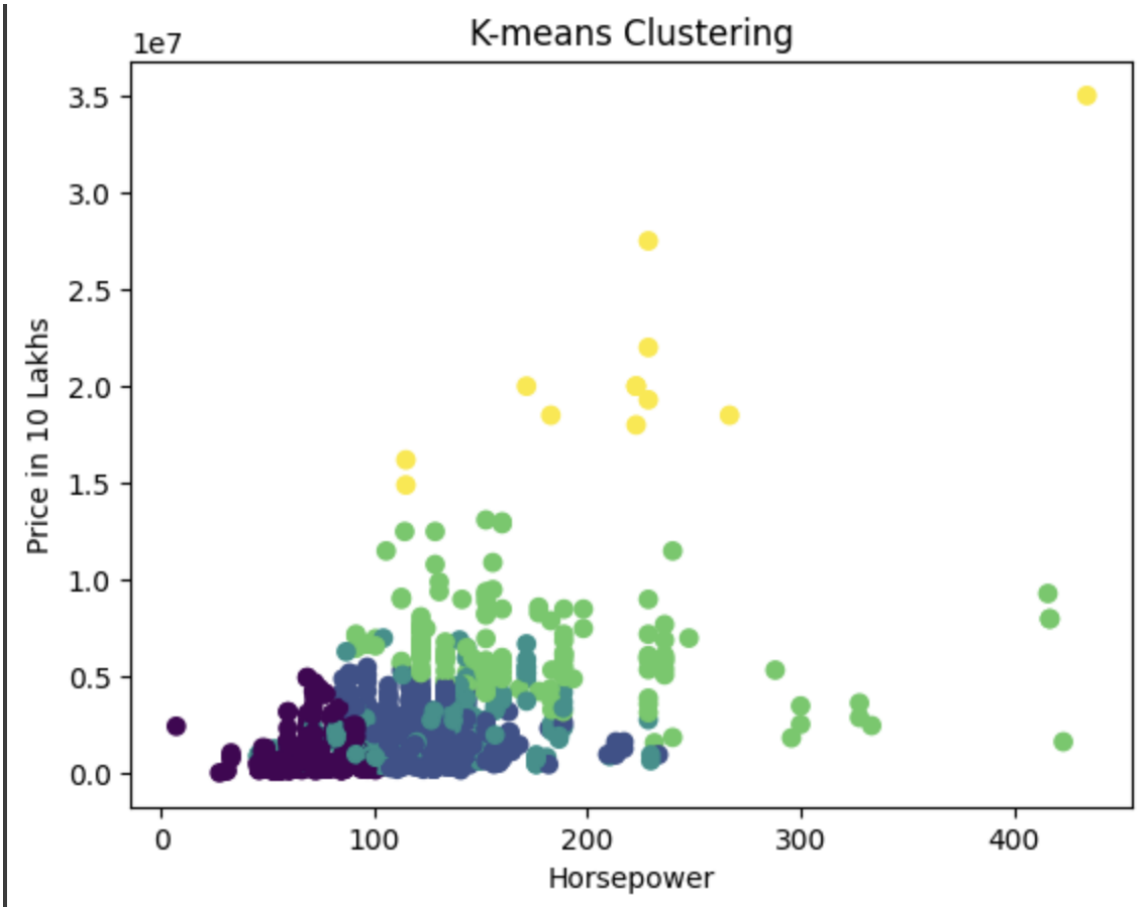
Upon examining the plot displaying the relationship between price and count of car models, a noteworthy observation emerges. The data indicates that the customer base is distributed relatively evenly across all price ranges, with only marginal variations in the number of customers. This finding suggests that the market for Electric Vehicles (EVs) encompasses a diverse range of price preferences, with no significant skew towards any particular price segment. Such market dynamics imply the importance of catering to customers across various price points when formulating effective market segmentation strategies for the EV market.

Segment Extraction :

By conducting K-means clustering on the provided dataset with three clusters and generating a scatter plot that visualises the relationship between price (in 10 Lakhs) and seating capacity, several insightful findings emerge. Notably, when examining the price range below 15 lakhs, a considerable proportion of customers display a preference for 7-seater cars. However, as the price range surpasses 15 lakhs, the majority of customers exhibit a preference for 5-seater cars, with a smaller percentage opting for 4-seaters. This suggests a shift in customer preferences based on price point, with higher-priced vehicles often featuring a seating capacity of 5 and lower-priced options accommodating larger groups with 7-seater configurations. These insights emphasise the importance of considering both price range and seating capacity as crucial

factors in market segmentation for the Electric Vehicle (EV) market. Additionally, further exploration and analysis of other variables in the dataset may provide additional valuable insights for effective segmentation strategies.



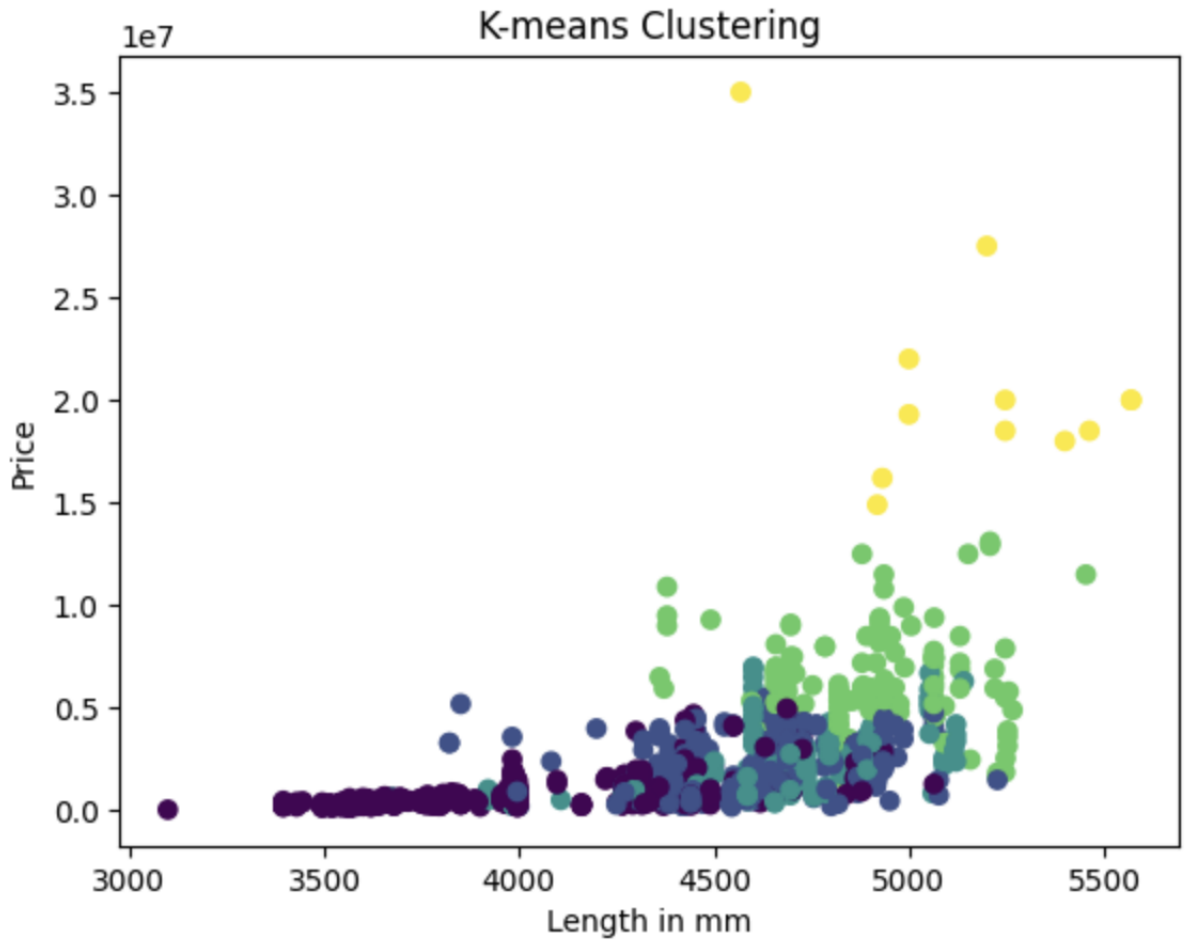


By visualising the scatter plot of price versus horsepower for cars in the dataset, a distinct pattern emerges, revealing the formation of four clusters. This observation provides valuable insights into customer preferences based on price and horsepower considerations.

In the price range below 10 lakhs, two distinct customer segments become evident. One segment prioritises features and is willing to compromise on horsepower, while the other segment values power and safety features. This differentiation highlights the diverse preferences and trade-offs made by customers in this price range.

Conversely, for cars priced above 10 lakhs, the majority of vehicles exhibit a horsepower rating exceeding 200. However, a minority of cars within this range possess horsepower below 200. This disparity suggests the existence of a niche market segment seeking lower horsepower options despite the higher price bracket.

These insights underscore the significance of understanding customer preferences in relation to both price and horsepower when devising effective market segmentation strategies for the Electric Vehicle (EV) market.



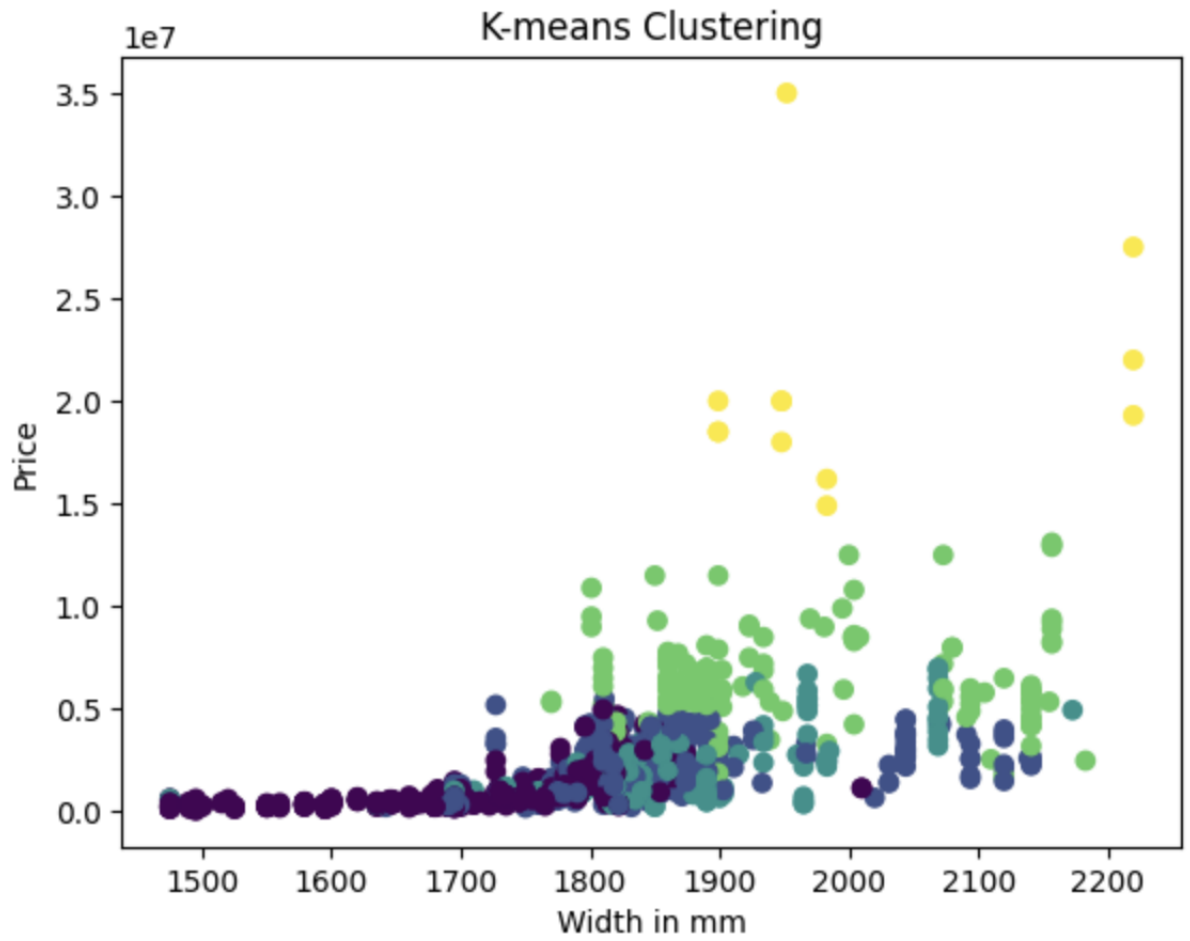
Upon plotting a scatter plot of car length against its price, several noteworthy insights are revealed.

Firstly, for cars priced below 5 lakhs, it is evident that the majority of vehicles fall below 4 metres in length. This observation highlights a significant segment of customers who prioritise

affordability and compactness in their vehicle choices. This price-driven segment values smaller cars that are well-suited for urban environments and offer cost-effective transportation solutions.

Furthermore, for cars priced above 5 lakhs, a distinct pattern emerges where the length of the cars ranges between 4.5 to 5.5 metres. This finding indicates another common customer preference related to car length in this price bracket. Customers in this segment tend to gravitate towards slightly larger vehicles, potentially seeking enhanced comfort, additional features, or greater interior space.

These insights shed light on the influence of price on customer preferences for car length, with distinct segments emerging based on affordability and desired vehicle dimensions. Understanding and leveraging these customer habituations related to car length can inform effective market segmentation strategies, allowing for tailored offerings that cater to different customer preferences and budgetary considerations.



Analysing the scatter plot of car width against its price reveals distinctive patterns and valuable insights for market segmentation.

In the price range below 5 lakhs, the width of cars spans a considerable range of 1500 to 1900 mm. However, the majority of cars in this price range tend to concentrate within the width range of 1800 to 1900 mm. This observation suggests that customers in this segment prioritise wider cars, potentially seeking increased interior space or a sense of stability on the road.

Conversely, for cars priced above 5 lakhs, the majority of vehicles fall within the width range of 1850 to 2000 mm. Additionally, a significant number of cars exceed 2000 mm in width. This finding indicates a separate market segment that values broader cars, possibly associating wider widths with enhanced comfort, a luxurious feel, or the desire for a more prominent presence on the road.

These insights highlight the impact of price on customer preferences for car width, with distinct segments emerging based on the desired width of the vehicles. Understanding and leveraging these customer preferences related to car width can inform effective market segmentation strategies, enabling the development of targeted offerings that cater to different customer needs and expectations.

Pratim Maity :

Packages and tools used:

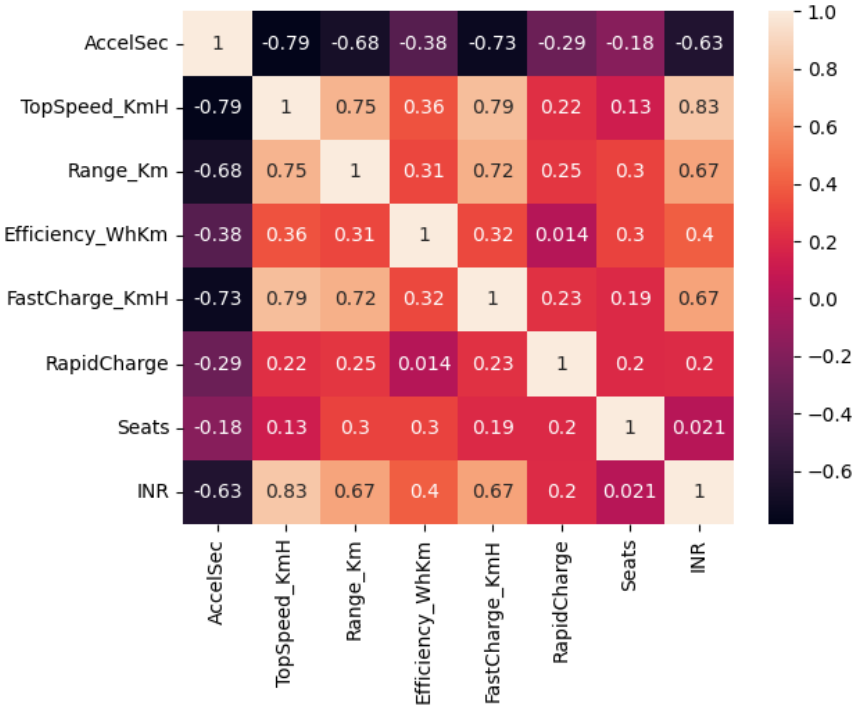
1. NumPy
2. Pandas
3. Scikit Learn
4. Seaborn
5. Matplotlib

Data Pre-processing:

Exploratory Data Analysis (EDA):

We started EDA with some data. In this we have compared our data in different aspects such as cars , top speed , price etc.

Correlation Matrix: It is a matrix which simply shows the correlation between each other. The relationship between two variables is usually considered strong when their correlation coefficient value is larger than 0.7.

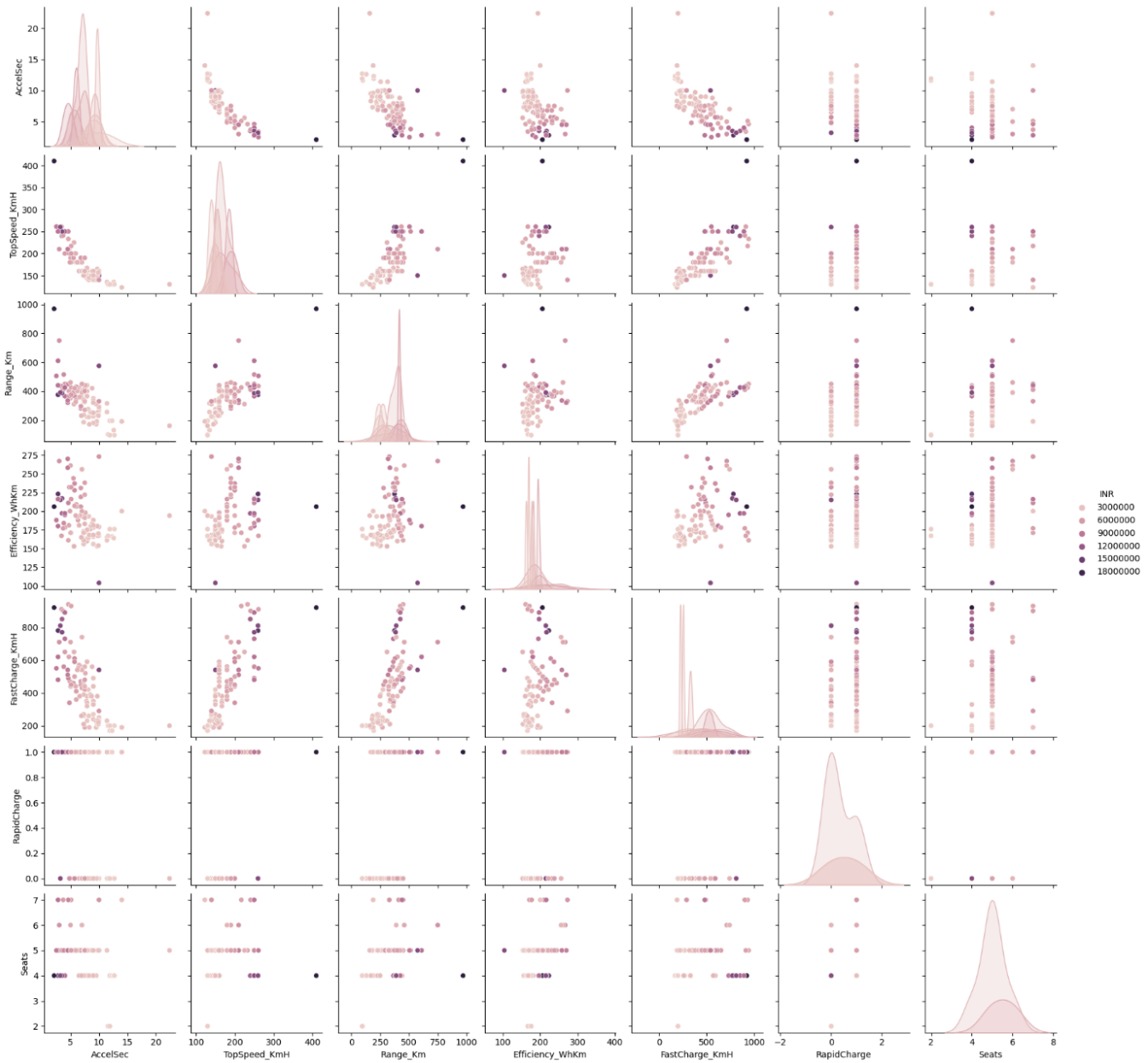


A pairplot is a type of data visualisation that allows you to explore the relationship between multiple variables in a dataset.

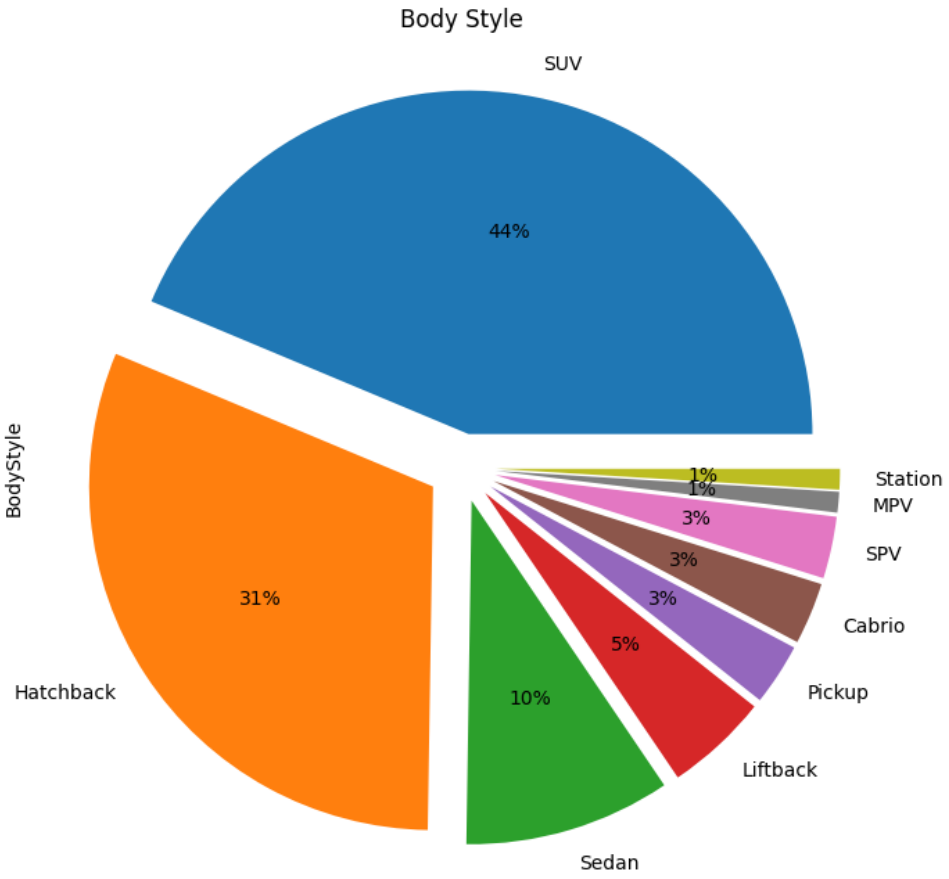
Pairplot to find how cost influences different aspects of a car

The diagonal subplots provide histogram or kernel density plots for each feature. These plots display the distributions of the variables and allow us to understand the range and frequency of values for each feature.

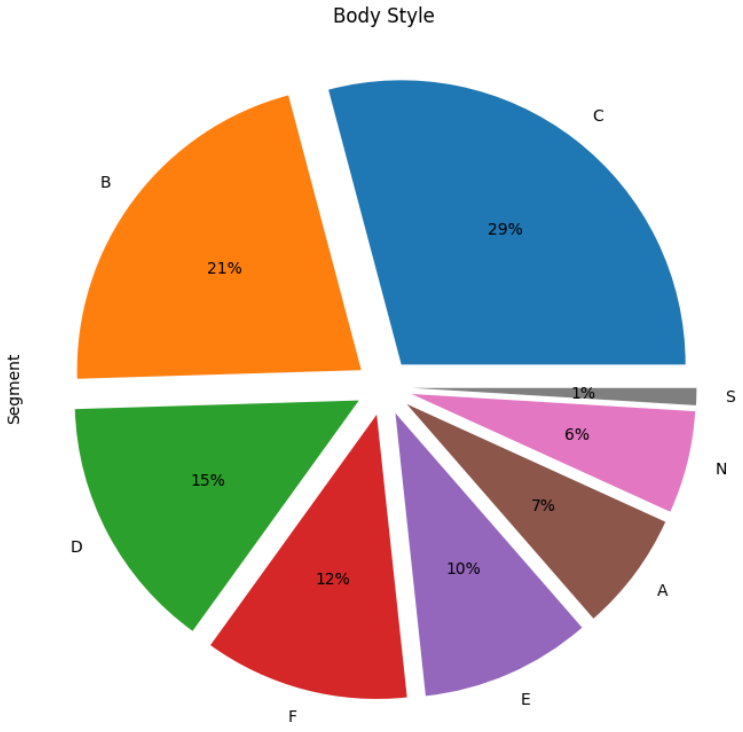
The off-diagonal subplots, we can observe the relationship between the price and each individual feature.



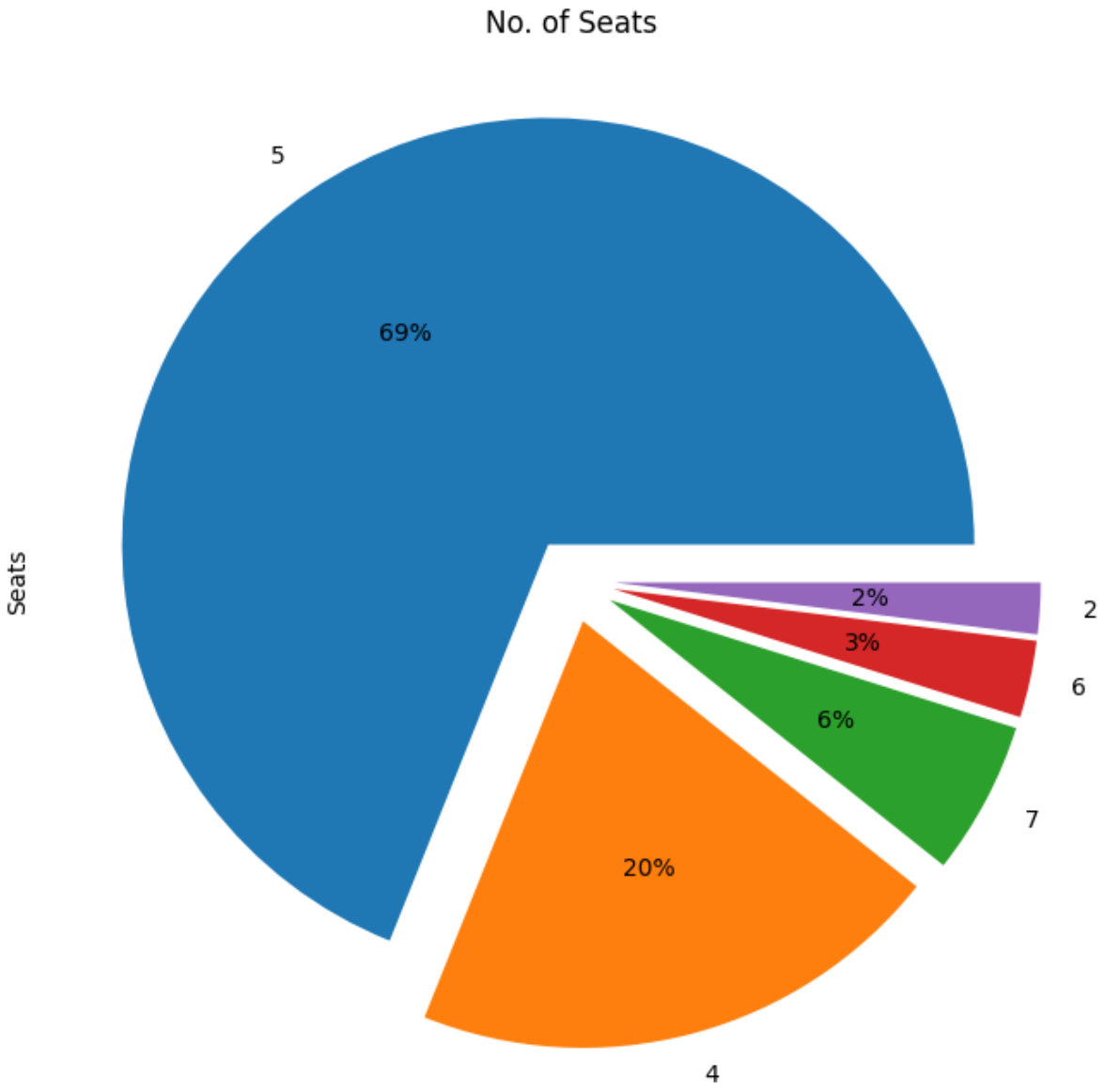
We have plotted the requirements of what type of car are most needed by the customer.



The pie plot in above represents that SUV is the highest preffered Body Style of a car by the customers.



The pieplot in above represents the segment which is the most manufactured and sold and also preferred by the customers.

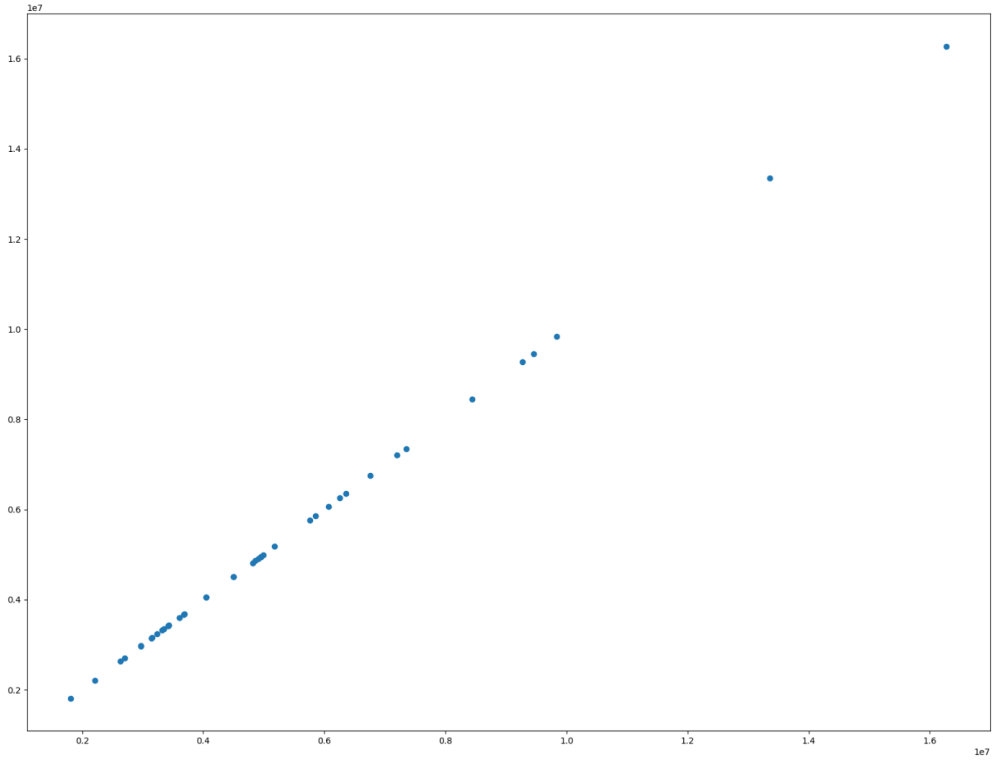


The pieplot above represents the most preferred number of seats among the customers.

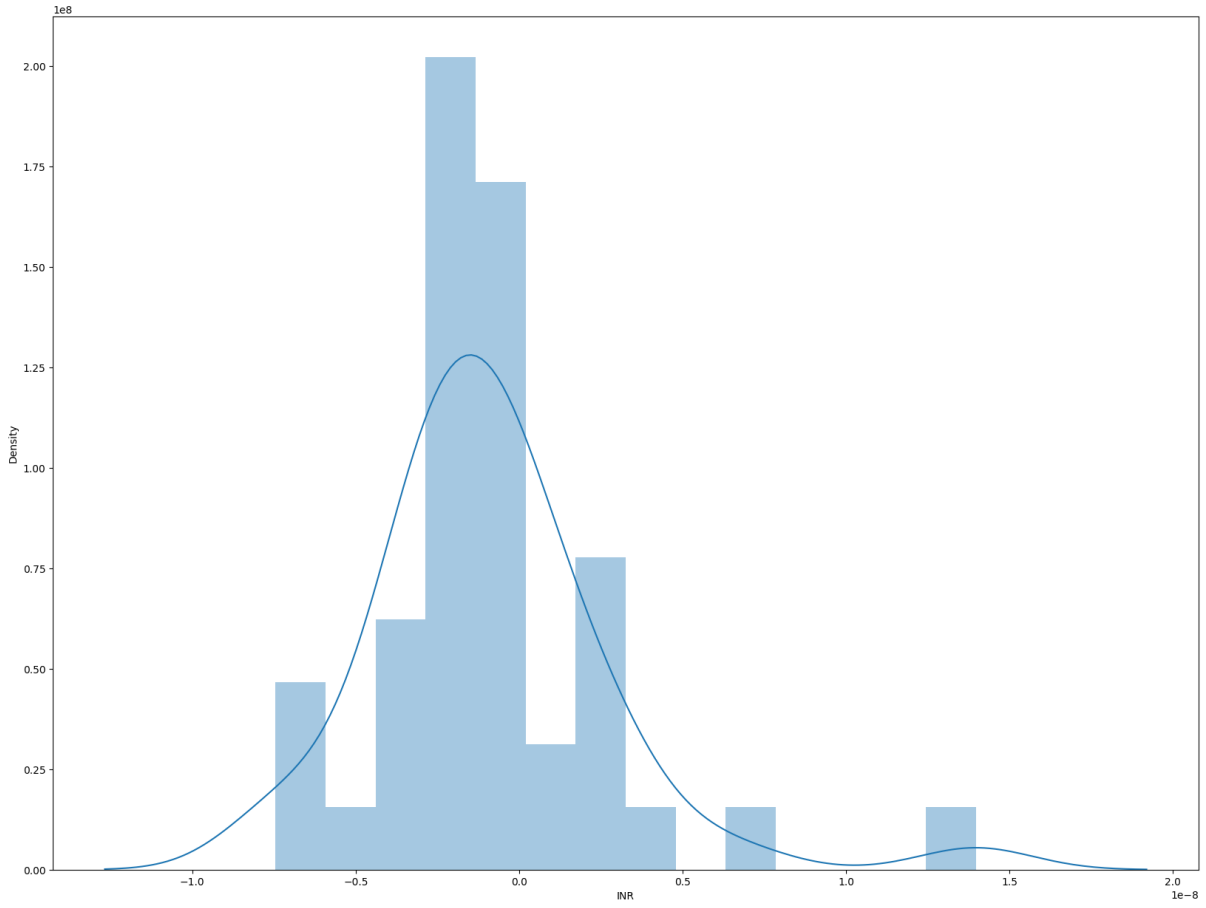
Prediction of Prices of most used cars

Linear regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models target prediction values based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Here we use a linear regression model to predict the prices of different cars from different brands.

After we train the Linear Regression Model, we test the remaining 40% of data on the model. The obtained results are checked using a scatter plot between predicted values and the original test data set for the dependent variable and acquired similarly to a straight line as seen in the figure below.



We have also plotted the density function which is normally distributed.



Conclusion :

It is quite evident from the data that the electric vehicle industry is still very small in comparison to the larger players already in the market. As is usual for late entrants into the market, carving a niche for a new product is paramount to performing well.

Segmentation of the Indian automobile market dataset revealed that the ideal segment to market to is between 35 to 40 years of age, and an annual total income of between 25 and 40 lakh rupees. This segment has a large enough number of dependents to warrant a larger vehicle, and purchasing power comparable to older segments, while being open to electric vehicles.

Further the market segmentation analysis revealed that customers display a strong preference for 5-seater cars, particularly in the price range below 15 lakhs. However, for higher-priced vehicles, the majority opt for 5-seaters while a significant percentage choose 4-seaters. Additionally, customers in the price range below 5 lakhs prioritise cars below 4 metres in length, while those in the higher price range prefer cars between 4.5 to 5.5 metres in length. There is also a correlation between price and width, with customers in the higher price range favouring wider cars. Furthermore, horsepower preferences vary, with a majority seeking vehicles with over 200 horsepower in the higher price range. These insights emphasise the need to consider factors such as seating capacity, price, length, width, and horsepower for effective market segmentation and targeted marketing strategies in the EV market in India.

Moreover we can conclude that most EVs have a power rating of 15 kW, indicating a preference for higher power capacity. Regional variations show that the NDMC region has a concentration of "DC-001" type EVs and 15 kW power-rated vehicles, while the Noida Authority region has a notable presence of "2 AC" type EVs and 142 kW power-rated vehicles. These findings highlight market dynamics, regional preferences, and potential market segments. Stakeholders can leverage these insights for targeted strategies and decision-making to drive EV adoption and market growth.

In conclusion, the analysis of the provided EV dataset reveals preferences for specific EV types and power ratings in different regions. Understanding these insights can inform decision-making processes for manufacturers, policymakers, and consumers, enabling targeted marketing, infrastructure planning, and product development to drive the growth and adoption of electric vehicles.